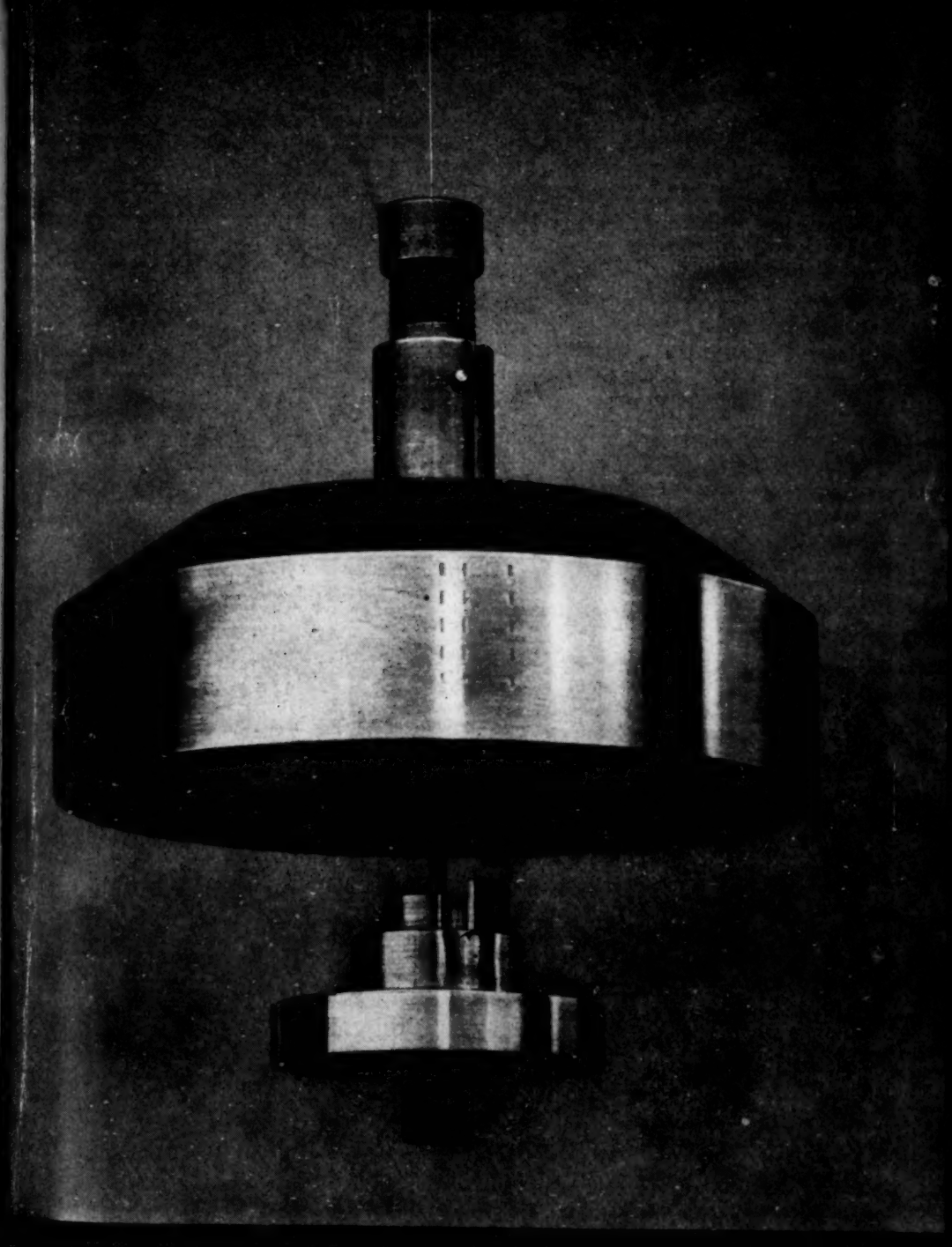
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 24, No. 6 JUNE, 1953

In This Issue

Rheology Symposium

		Page
Annual Meeting of the Society of Rheology at the Franklin Institute, Philadelphia, October, 1952		665
Effects of Nuclear Radiations on the Mechanical Properties of Solids	G. J. Dienes	666
Test of the Theory of Secondary Viscoelastic Stress	M. Mooney	675
Dynamic Mechanical Properties of the System Polystyrene-Decalin	Lester D. Grandine, Jr., and John D. Ferry	679
Mechanical Investigations of Elastomers in a Wide Range of Frequencies	Wladimir Philippoff	685
Viscoelastic Properties of Dilute Polymer Solutions	P. E. Rouse, Jr., and K. Sittel	690
Temperature Coefficients of Non-Newtonian Viscosity at Fixed Shearing Stress and at Fixed Rate of Shear	A. B. Bestul and H. V. Belcher	696
Effects of Rate of Shear on Inherent and Intrinsic Viscosities of Polystyrene Fractions	L. J. Sharman, R. H. Sones, and L. H. Cragg	703
Anomalous Viscous Flow at Very Low Rate of Shear and Small Shearing Stress	J. T. Bergen and W. Patterson, Jr.	712

Regular Contributed Original Research

Signal-to-Noise Ratios in Band-Pass Limiters	W. B. Davenport, Jr.	720
Space Charge Requirements in Some Ideally Focused Electron-Optical Systems	Nelson Wax	727
Investigation of Boron Carbide	Frank W. Glaser, David Moskowitz, and Benjamin Post	731
Creep of Zinc Single Crystals under Direct Shear	Elliot H. Weinberg	734
Tolerance Coefficients for *R-C* Networks	Charles Belove	745
Simplification for Mutual Impedance of Certain Antennas	Jesse Gerald Chaney	747
Mutual Impedance of Rhombic Antennas Spaced in Tandem	Jesse Gerald Chaney	751
Slot Coupling of Rectangular and Spherical Wave Guides	L. B. Felsen and N. Marcuvitz	755
Determination of the Drag on a Cylinder at Low Reynolds Numbers	R. K. Finn	771
Charged Disk in Cylindrical Box	W. R. Smythe	773
Darlington Problem	Louis Weinberg	776
Activation Energy for Viscous Flow and Short-Range Order	G. J. Dienes	779
Output Signal-to-Noise Ratio of a Power-Law Device	Nelson M. Blachman	783
Electromagnetic Transmission Characteristics of the Two-Dimensional Lattice Medium	Herbert S. Bennett	785
Letters to the Editor:		
Pure Gliding of Metal Crystals	Hiroshi Kanzaki	811
Analysis of Fluorescent X-Radiation by Means of Proportional Counters	Robert E. Rowland	811
Dynamic Compliance, Dynamic Modulus, and Equivalent Voigt and Maxwell Models for Polyisobutylene	Herbert Leaderman and Robert S. Marvin	812
X-Ray Line Broadening from Precipitation in Cu-Fe Alloys	T. S. Hutchison	813
Use of the Mercury Cathode in the Dielectric Breakdown of Liquids	A. H. Sharbaugh, J. K. Bragg, and R. W. Crowe	814
Note on "Metallic Shadow-Casting Using a Nozzle System"	Shigeto Yamaguchi	815
Expansion of a Fissionable Gas as a Source of Atomic Power	Edward H. Kerner	815
Nonreciprocal Microwave Component	M. L. Kales, H. N. Chait, and N. G. Sakiotis	816
Errata:		
Emission Decay Phenomena due to the Contaminated Anode	K. Amakasu and T. Imai	817
Announcements		818
Index to Advertisers		xxii

ELMER HUTCHISSON, *Editor*

EARLE C. GREGG, *Assistant Editor*

Board of Editors

RAY E. BOLZ
LYLE D. BORST
WALLACE J. ECKERT
ROBERT D. HEIDENREICH
LEONARD B. LOEB
W. JAMES LYONS
LOUIS A. PIPES
ROMAN SMOLUCHOWSKI
W. R. SMYTHE

Publication of the American Institute of Physics

OFFICERS

GEORGE R. HARRISON, *Chairman*
GEORGE B. PEGRAM, *Treasurer*

GOVERNING BOARD

J. W. BUCHTA
KARL K. DARROW
G. J. DIENES
S. A. GOUDSMIT
GEORGE R. HARRISON
FREDERICK V. HUNT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
GEORGE B. PEGRAM
I. I. RABI
DUANE ROLLER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing

MELVIN LOOS

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Advertising Manager*
RUTH F. BRYANS, *Publication Manager*
BETTY H. GOODFRIEND, *Assistant Secretary*
EILEEN B. NEUBERGER, *Circulation Manager*
KATHRYN SETZE *Assistant Treasurer*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 24, Number 6 June, 1953

Annual Meeting of the Society of Rheology
at the Franklin Institute, Philadelphia, October, 1952

IN 1952 the Society of Rheology returned for its Annual Meeting to Philadelphia where for two days sessions were held under the auspices of the Franklin Institute. Seventeen papers were presented. Thanks to the cooperation of the American Institute of Physics, eight of these papers will be found in full in this issue of the *Journal of Applied Physics*. Thus it has been possible to continue the policy of the Society of Rheology to collect the available papers from its Annual Meetings into single "Rheology" issues of friendly host periodicals. At this juncture we wish to extend out thanks to the *Journal of Colloid Science* for a similar generous cooperation in the past.

The papers presented at the meeting may be arranged under the following headings, whereby those not appearing in this Journal will be given with their full titles and a brief note as to their contents.

The first of these general papers was given by R. D. Andrews and T. Alfrey, The Dow Chemical Company, Midland, Michigan: "Review of dynamic methods," and will be published in one of the next issues of this Journal. The other two papers were:

Finite Elastic Deformation. R. S. RIVLIN, *Naval Research Laboratory, Washington, D. C.*—A brief account will be given on the basis of the theory by which load-deformation relations for bodies of elastic materials may be calculated when the strains are not infinitesimally small. Some of the principal results will be given and it will be shown how these may be related to the measured load-deformation relations for vulcanized rubber.

Proposed Nomenclature and Symbolism for Linear Viscoelastic Behavior of High Polymers. H. LEADERMAN, *National Bureau of Standards, Washington, D. C.*—There is at present a lack of uniformity in the use of terms and symbols describing the observable and derived quantities associated with the linear viscoelastic behavior of high polymers. A logical system of nomenclature which is likely to be adopted must be modified to take account of convenience and of current usage.

The next session was opened with the paper by G. J. Dienes, and followed by two papers on styrene filaments by K. J. Cleereman and by H. J. Karam:

Physical Properties of Oriented Polystyrene Monofilament. K. J. CLEEREMAN, *Dow Chemical Company, Midland, Michigan.*—This paper discusses the fabricating variables that must be considered in the production of polystyrene monofilament. The effect of these fabricating variables on physical properties is also discussed. On the basis of this discussion a suggested method of fabricating highly oriented polystyrene monofilament is given.

Refraction Study of Oriented Polystyrene Monofilament. H. J. KARAM, *Dow Chemical Company, Midland, Michigan.*—Retraction of oriented polystyrene monofilament was studied. A method to obtain a retraction curve for oriented fibers is described. The portion of the retardation time spectrum oriented was calculated from the data. The effect of fabrication variables (percent hot stretch and stretching bath temperature) on the spectrum are described.

Mooney's paper on Stress Analysis led three papers on the Visco-Elastic Properties of Polymers as tested by dynamic methods—namely, those of L. D. Grandine and J. D. Ferry, W. Philippoff, and P. E. Rouse and K. Sittel.

The next three papers by A. B. Bestul, by J. T. Bergen and William Patterson, and by L. H. Cragg, L. J. Sharman, and R. H. Sones concerned themselves with the relation between shearing stress and rates of strain.

Of the final four papers referring to measuring and testing techniques, that by S. R. Rich and W. Roth will be found in an early subsequent issue. The other three are:

Effect of Surfaces of Rotor and Dies on Mooney Viscosity. G. E. DECKER AND FRANK L. ROTH, *National Bureau of Standards, Washington, D. C.*—It was found that most commercial rubbers adhere to the metal surfaces sufficiently well so that serration in the dies and rotors are not required to prevent slippage. In studies of synthetic rubbers of relatively high molecular weight, slippage of the rubber on the metal surfaces

causes erratic and low apparent viscosities and is a function of the type of metal in contact with the specimen. Decreasing the speed of the rotor eliminates these anomalous results.

Sonic Measurements on Grinding Wheels. W. M. FELTON, *Franklin Institute, Philadelphia, Pennsylvania.*—This paper describes a nondestructive method for testing whether the manufactured grade of a grinding wheel corresponds to the designed grade. This method includes a sonic testing device for measuring natural frequency and a mathematical expression for a grade index that is proportional to the velocity of sound.

Sonic Grade Comparators for Testing Grinding Wheels. J. E. PRICE, *Simonds Abrasive Company, Philadelphia, Pennsylvania.*—A discussion will be given of operational problems and design changes necessary to achieve proper utilization of a sonic comparator developed by FI for use in determination of grinding wheel grade variations. Possible use of sonic data for determination of cracked wheels is discussed.

F. R. EIRICH.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Effects of Nuclear Radiations on the Mechanical Properties of Solids*

G. J. DIENES
Brookhaven National Laboratory,† Upton, Long Island, New York

(Received Februrary 16, 1953)

The general nature of radiation effects in solids is reviewed briefly. Current theoretical understanding of the mechanical properties of solids is critically evaluated. The effect of nuclear radiation on the mechanical properties is discussed in detail. It is shown that the changes in the mechanical properties of crystalline substances (mostly metals) can be quite satisfactorily interpreted on the basis of the production of interstitial atoms and vacant lattice sites by fast particle irradiation. Isolated vacancies and interstitials may not be able to account for all the observations, and attention is called to the possible need of postulating the existence of aggregates of these lattice defects. In molecular solids (mostly high polymers) nuclear radiations bring about changes in the substance which are best described as chemical ones. Ions and free radicals are formed leading to subsequent chemical reactions thereby altering the properties of the substance. Drastic changes in the mechanical properties of high polymers are observed. Correlation with structural changes has hardly been started. Experiments are suggested which, in the writer's opinion, should give further insight into the fundamental processes involved.

I. INTRODUCTION

IT is well known by now that nuclear radiations may alter drastically the properties of solid materials. A considerable amount of fundamental and applied research has been carried out in this field during the last decade. While some aspects remain classified, the main experimental and theoretical features of these new phenomena have been described in the open literature during the last five years and an increasing number of papers deal with current research in the field.

Early work of the Manhattan Project has been described by Burton.[1,2] The main theoretical results have been given by Seitz.[3] Studies on semiconductors and metals have been reviewed by Lark-Horovitz[4] and Billington and Siegel;[5] chemical aspects have been treated by Burton.[6] The whole field of the effects of radiation on materials has been critically reviewed by Slater, from a theoretical as well as experimental point of view, in a report written during the summer of 1949 and recently published as a declassified article.[7] The progress of the last three years has recently been reviewed by the writer.[8] A brief historical account has been given by Seitz.[9]

The purpose of this paper is to focus attention on one set of physical characteristics of materials, namely the mechanical properties, and to discuss the effects of radiation on the first of three properties in more detail than is possible in a more general article. Much of the work in this field until recently has been concerned mainly with the accumulation of practical information. In the last few years, however, fundamental principles and generalization began to emerge and attention will be focused on these in this paper.

In Section II a brief account is given of the nature of radiation effects. In Section III the effects of nuclear radiation on mechanical properties of solids are discussed in detail, and in Section IV some experiments are suggested which, the writer feels, are of importance in giving further insight into the fundamental processes involved.

II. NATURE OF RADIATION EFFECTS

Only a very brief discussion of the nature of radiation effects will be presented here since detailed treatments

* Presented at the Annual Meeting of the Society of Rheology, Franklin Institute, October 30–31, 1952.

† Research carried out under contract with the U. S. Atomic Energy Commission.

[1] M. Burton, J. Phys. & Colloid Chem. **51**, 611 (1947).
[2] M. Burton, Chem. Eng. News **26**, 1764 (1948).
[3] F. Seitz, Disc. Faraday Soc. No. 5, 271 (1949).
[4] K. Lark-Horovitz, "Nucleon bombarded semi-conductors," *Reading Conference on Semi-Conducting Materials* (Butterworth's Scientific Publications, London, 1951), pp. 47–78.
[5] D. S. Billington and S. Siegel, Metal Prog. **58**, 847 (1950).
[6] M. Burton, Chapter on Radiation Chemistry in *Annual Review of Physical Chemistry* (Annual Reviews, Inc., Stanford, 1950), Vol. I., pp. 113–132.
[7] J. C. Slater, J. Appl. Phys. **22**, 237 (1951).
[8] G. J. Dienes, "Radiation effects in solids," Chapter in *Annual Review of Nuclear Science* (Annual Reviews, Inc., Stanford, 1953), Vol. II., pp. 187–221.
[9] F. Seitz, Physics Today **5**, No. 6, 6 (1952).

are available in the literature.[3,6–8] Fast neutrons, upon collision with atoms of any material, produce charged recoil atoms which in turn produce secondary recoils and so on. The major changes produced in a material result from the action of the fast charged recoil atoms which are far more numerous than the incident neutrons. These recoil atoms are responsible for three prominent changes in a solid.

(1) Ionization effects. A large fraction of the energy of a fast charged particle is dissipated in ionization and excitation as the particle is slowed down. In a good conductor the ionization effects will disappear very quickly and will only contribute to the heating of the material. In an insulator the electrons liberated by ionization may be trapped at various lattice imperfections resulting in more or less permanent changes in the substance. In a molecular compound ionization brings about changes which are best described as chemical ones. Ions and free radicals are formed which can lead to chemical reactions thereby altering the original substance.

(2) Displacement effects. In a crystal, atoms will be knocked out of their lattice positions leaving vacant lattice sites behind and finally coming to rest in interstitial positions. It is evident that if an appreciable number of displaced atoms are created many of the physical properties of the substance may be changed drastically. Many of the experimentally observed phenomena in crystals are well explained and correlated on the basis of the vacancy interstitial picture. It should be pointed out, however, that these disturbances may be highly localized in clusters along the track of the primary particle so that, in some cases at least, it may be more profitable to think about a disordering of the lattice akin to that produced by cold work or rapid local heating and quenching (thermal spike).

(3) Thermal spikes. A thermodynamic or statistical approach to damage production is also possible. Dissipation of the energy of the fast particle may be viewed as the introduction of a large amount of heat at a localized point in the material, and a region of material around the track of the particle will be heated to a high temperature. The region of excitation will rapidly expand accompanied by a fast fall in temperature. The result is a so-called "thermal spike," i.e., rapid heating and quenching of a small volume of the material. A fraction of the atoms will be left in displaced positions after the cooling of a spike, and the end result is equivalent to a localized production of vacancies and interstitials. In this picture, however, the energy transmitted to the atoms without actually displacing them is also taken into account. This is essentially thermal energy which may stimulate any of the effects which high temperature can bring about. The best known example is the disordering of ordered alloys by irradiation. The notion of thermal spikes has been useful in interpreting such phenomena.

In this paper attention will be focused on displacement and ionization effects. Most of the work to date on the effects of radiation on mechanical properties has been done on metallic and high polymeric systems. As indicated above ionization effects are not expected to be of any importance in metals and the phenomena will be interpreted on the basis of the presence of interstitials and vacancies in the crystalline lattice. In the case of high polymers ionization and subsequent chemical changes are the important effects. In case of pile bombardment much of the ionization is caused by fast moving hydrogen ions ejected by elastic collision with fast neutrons. In general, all types of high energy ionizing radiation (α, β, γ, fast neutrons, etc.) lead to about the same changes in the material. In detailed radiation chemistry studies, of course, differences in ionization densities must be considered.

III. EFFECTS OF RADIATION ON MECHANICAL PROPERTIES

1. Theoretical Understanding of the Mechanical Behavior of Solids

It is probably no exaggeration to state that the least understood, and probably the most important, physical characteristics of solids are their mechanical properties. Before radiation effects, which are known to change these properties extensively in many cases, are discussed in detail, it is pertinent to review this field critically. Some mechanical properties are far more suitable than others for a fundamental study of radiation effects, because of the status of knowledge in the absence of radiation.

Theoretical understanding of the mechanical properties of materials may be discussed at two levels: the phenomenological and the atomic or molecular. Phenomenological understanding means that a group of phenomena can be quantitatively correlated and described. Atomic or molecular understanding goes considerably further and requires that a group of phenomena be describable and predictable from known atomic or molecular characteristics.

The above remarks are best illustrated by a discussion of "perfect elasticity." One may anticipate at this point by stating that only within the range of perfect elasticity are the elastic properties of solids understood both on a phenomenological and atomistic basis.

A material is said to be perfectly elastic if:

1. Stress and strain are proportional to each other (Hooke's law).
2. Strain is completely and instanteneously recoverable upon release of load.
3. Stress and strain are independent of time—i.e., stress is only a function of strain, $\sigma=\phi(\epsilon)$.

The response of such a material under any sequence of stresses and strains is describable by the theory of elasticity, at least in principle, although in practice one may be limited to small strains. Within this range the material is completely characterized by a set of

elastic moduli. Thus, within the elastic range of behavior a satisfactory phenomenological theory exists.

At the atomic or molecular level, the elastic constants of metals[10–12] and ionic crystals[13,14] are theoretically calculable from the interatomic forces, and in the few cases where the calculations have been carried out the results are in good agreement with experiment. The theoretical calculations are difficult even for simple solids. It is probably correct to say, however, that the theoretical difficulties are difficulties of technique and not of principle, i.e., it is known how the problem is to be formulated although the detailed calculations may be very difficult to carry out. In this sense, then, one can claim theoretical understanding at both levels within the range of perfect elasticity.

In a similar way one may sum up the present status of the theoretical understanding of other mechanical properties. The next step in complexity involves time-dependent phenomena. Stress relaxation, creep and internal friction at very low stresses and strains belong to the class of anelastic or viscoelastic phenomena.‡ Within this range of mechanical behavior the properties of the material are completely describable by means of linear differential equations relating stress, strain, and their time derivatives. Stress and strain are still linearly related and the material obeys the Botzmann superposition principle.[15,16] It may be remarked, however, that the stress-strain curve of an anelastic or viscoelastic material is never linear at any finite rate of stressing and straining,[17] and it is important to differentiate clearly between the nonlinearity of stress-strain curves (in shear) due to time effects and that due to an inherent nonlinearity of stress to strain. Within the anelastic range of mechanical behavior a satisfactory phenomenological theory exists.

At the atomic or molecular level our knowledge is far less complete. The field is well summarized by Zener[15] for metals, by Alfrey[16] for high polymers and Leaderman[18] for textiles and other materials. Anelasticity phenomena in metals have been used successfully for studying various metallurgical processes. Two important conclusions should be mentioned:

1. Relaxation across grain boundaries is an important source of internal friction, and changes in grain boundary structure are conveniently followed by means of such measurements.

2. Impurities, in terms of interstitial or substitutional atoms, are another important cause of internal friction. The atoms in a disordered solid solution of cubic structure are distributed isotropically. In the presence of an applied stress, however, a preferential equilibrium distribution may be induced. The continual striving of the metal, via microscopic diffusion (local atomic exchanges), to maintain an equilibrium distribution gives rise to anelasticity. Recent researches of Kê,[19] Zener,[20] Wert,[21] Nowick,[22] and LeClaire[23] have shown the usefulness of this technique in studying the properties, and particularly the diffusion, of imperfections in crystals.

In the high polymer field, a large number of workers[24] have been investigating the relation of viscoelastic properties to chemical structure, molecular configuration, and to bond breakage and bond formation. Full understanding can by no means be claimed although considerable progress has been made in the last few years. Similarly, anelastic properties of glasses are beginning to be understood in a more definite manner.[25]

In conclusion one may state, therefore, that some anelastic and viscoelastic phenomena are quite well understood at the atomic and molecular level.

As the stress is increased beyond the anelastic range the plastic properties of solids become predominant. The essential characteristics of the plastic range of behavior is the inherent nonlinearity of shear§[26,27] stress to shear strain coupled with relatively large permanent deformations.

Time-dependent phenomena (creep, stress-relaxation) in the plastic range may be quite similar to viscoelastic behavior with respect to the time parameter but not with respect to response to stress. In this range a small change in stress results in a much larger than proportional change in strain. This is the case apart from the time effect, i.e., the deformation under a constant stress at a fixed time is not proportional to the applied stress in the plastic range. In many cases, the conse-

[10] J. Bardeen, J. Chem. Phys. **6**, 367 (1938).
[11] K. Fuchs, Proc. Roy. Soc. (London) **A153**, 622 (1936); **A157**, 444 (1936).
[12] D. S. Leigh, Phil. Mag. **42**, 139 (1951).
[13] R. Landshoff, Z. Physik **102**, 201 (1936).
[14] P. O. Löwdin, "A theoretical investigation into some properties of ionic crystals," thesis, Uppsala, (1948).
‡ The term anelastic (reference 15) has been used primarily with reference to metals, while the term viscoelastic (reference 16) has been generally employed in discussing organic materials (particularly high polymers) and glasses.
[15] C. Zener, *Elasticity and Anelasticity of Metals* (The University of Chicago Press, Chicago, 1948).
[16] T. Alfrey, *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948).
[17] G. J. Dienes, J. Colloid Sci. **2**, 131 (1947).
[18] H. Leaderman, "Elastic and creep properties of filamentous materials and other high polymers" (The Textile Foundation, Washington, D. C., 1943).
[19] T. S. Kê, Phys. Rev. **74**, 9, 16 (1948); J. Appl. Phys. **19**, 285 (1948); Am. Inst. Mining Met. Engrs., Metals Technol. **15**, 2370 (1948).
[20] C. Zener, Trans. Am. Soc. Mech. Engrs. **152**, 122 (1943); Phys. Rev. **71**, 34 (1947).
[21] C. A. Wert, Phys. Rev. **79**, 601, (1950); J. Appl. Phys. **21**, 1196 (1950).
[22] A. S. Nowick, Phys. Rev. **82**, 551 (1951).
[23] A. D. LeClaire, Phil Mag. **42**, 673 (1951).
[24] For bibliography of recent research see R. Simha, *et al.* "Bibliography of recent research in the field of high polymers," Natl. Bur. Standards Circular No. 498, (1950), pp. 1–9, 14–17.
[25] See, for example, J. V. Fitzgerald, J. Am. Ceram. Soc. **34**, 314, 339, 388, 390 (1951).
§ This restriction to shear is of importance. Rubber shows a linear "equilibrium" stress-strain curve in shear up to large strains. The relation in tension or compression is highly nonlinear because of the geometry of large strains (references 26, 27).
[26] M. Mooney, J. Appl. Phys. **11**, 582 (1940); **19**, 434 (1948).
[27] R. S. Rivlin, Trans. Roy. Soc. (London) **A240**, 459 (1948).

TABLE I. Theoretical understanding of the mechanical behavior of solids.

Mechanical behavior	Crystalline materials		Amorphous materials	
	Phenomenological	Atomic or molecular	Phenomenological	Atomic or molecular
Perfect elasticity	Yes	Yes	Yes	Yes
Anelasticity (viscoelasticity)	Yes	Some phenomena: yes Generally: no	Yes	Partial
Plastic properties	Partial	Partial	No	No
Fracture and strength	No	No	No	No

quence of large applied stresses (or strains) is a more or less permanent change of the properties of the material, usually referred to as strain or work hardening. Further, for many materials a fairly well-defined yield stress may be measured which separates the plastic region from the elastic one. The yield stress is much lower than predicted theoretically from the force necessary to cause plastic deformation (slip, glide) in a perfect crystal.

In order to explain the low yield stresses of crystalline materials, specific types of lattice imperfections, called dislocations have been proposed by Taylor,[28] Polanyi,[29] and Orowan[30] based on earlier suggestions made by Prandtl[31] and Dehlinger.[32] An elaborate theory of dislocations has been developed since then. For critical discussions of this theory the reader is referred to several excellent reviews.[33–38]

The status of the theory appears to be as follows. In the last few years dislocation theory has begun to correlate a number of diverse structure sensitive mechanical properties. This theory is complex and reliable quantitative calculations are extremely difficult to make. Qualitatively, work hardening, critical shear stress, and even some aspects of creep and internal friction appear to be explicable on the basis of this theory. The predictive usefulness of the theory, particularly in a quantitative manner, is still very limited. At the phenomenological level, therefore, one can claim partial understanding. At the atomic level, it appears now that reasonable hypotheses can be formulated as to the origin, creation, and multiplication of dislocations.[33,39] Thus, a partial understanding can also be claimed here.

The situation is certainly not better in the case of amorphous materials. In a very qualitative way, one knows that plastic properties in these materials depend on the movement of molecules, or large groups of molecules, relative to each other. Orientation and crystallization phenomena apparently play an important role in determining the plastic properties. Phenomenological descriptions have been achieved in some cases,[40–43] but one can hardly claim understanding at either level.

Rubber falls somewhat into a class of its own, since it behaves neither as a plastic material nor as a simple linear viscoelastic material. The statistical theory of rubber elasticity describes quite successfully the properties of this interesting substance. The theory is essentially based on the analysis of the configurational response of a network consisting of a series of mesh points connected by flexible chains of varying length. The theory provides a rather high level of understanding at both the phenomenological and molecular level.[16,24]

Fracture and strength are probably the least understood mechanical properties. The subject is critically discussed in several reviews.[16,44,45] At the phenomenological level the following points need to be emphasized: (a) it is not all certain that strength can be described in terms of a true material constant; (b) fracture may not be a single physical phenomenon in the sense that many different processes may lead to disintegration; and (c) there is no agreement among workers in the field as to the criterion to be used for describing fracture (critical stress, strain, energy, etc.). A great deal of theoretical work has also been done on the problem of crystal stability. From several different viewpoints the result is the same; the theoretical breaking stress, as compared to experiment, is too high by a factor of about 10^3. Note the similar discrepancy between theoretical and experimental yield stress discussed above. It is precisely this discrepancy which led to theories of imperfections, specifically that of dislocations for crystalline materials. The theoretical calculations have been questioned by Fürth,[46] and more recently by Saibel,[47] on the grounds

[28] G. I. Taylor, Proc. Roy. Soc. (London) **145**, 362 (1934).
[29] M. Polanyi, Z. Physik **89**, 660 (1934).
[30] E. Orowan, Z. Physik **89**, 634 (1934).
[31] L. Prandtl, Z. angew. Math. Mech. **8**, 85 (1928).
[32] U. Dehlinger, Ann. Physik **2**, 749 (1929).
[33] F. Seitz, *Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943); *Advances in Physics*, Phil Mag. Supplement **1**, 43–90 (1952).
[34] "Report of a conference on the strength of solids" University of Bristol, England, [Physical Society, London (1948)].
[35] Pittsburgh Symposium on the Plastic Deformation of Crystalline Solids (U. S. Office of Naval Research, 1950).
[36] A. H. Cottrell, *Progress in Metal Physics* (Interscience Publishers, Inc., New York, 1949).
[37] F. R. N. Nabarro, *Advances in Physics*, Phil. Mag. Supplement **1**, 271–376 (1952).
[38] F. C. Frank, Phil Mag. **43**, 809 (1951).
[39] F. C. Frank, *Advances in Physics*, Phil. Mag. Supplement **1**, 91–109 (1952).
[40] Halsey, White, and Eyring, Textile Research J. **15**, 295 (1945); **15**, 451 (1945), and subsequent papers.
[41] Alexander Brown, J. Appl. Phys. **20**, 552 (1949).
[42] Hopkins, Baker, and Howard, J. Appl. Phys. **21**, 206 (1950).
[43] Carey, Dienes, and Schulz, Ind. Eng. Chem. **42**, 842 (1950).
[44] M. Gensamer, *et al.* "The fracture of metals" (American Welding Society, New York, 1947).
[45] E. Orowan, "Fracture and strength of solids" (Cavendish Laboratory, Cambridge, NP-1077, November 15, 1949).
[46] R. Fürth, Proc. Roy. Soc. (London) **177A**, 217 (1941).

that an improper criterion of breakdown has been used. These workers propose a "melting theory" of strength (based on some earlier work of Born[48]); that is, their criterion for breakdown is that the elastic energy per unit volume be equal to the heat needed to melt the same amount of substance. This postulate leads to theoretical values in much better agreement with experiment. Considerable further development is necessary before the merits of the various hypotheses can be properly judged.

This evaluation of our understanding of the mechanical properties of substances is summarized in Table I. Such a table is useful as a framework for discussing the effects of nuclear radiation and for designing fundamental experiments.

2. Radiation Effects in Crystalline Solids—Mostly Metals

As already stated in Section II, the most important effects of nuclear radiation on crystalline substances arise from a disordering of the lattice via the production

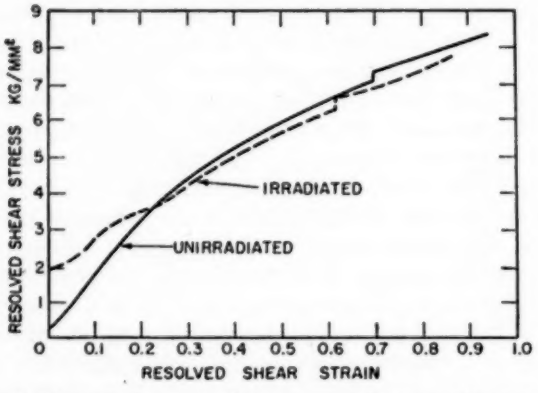


FIG. 1. Stress-strain curve of copper crystals (from reference 53).

of displaced atoms. In this section attention will be focused on metal crystals since practically no work has been reported in the unclassified literature on changes in the mechanical properties of nonmetallic crystals. It is known, of course, that displaced atoms play an important part in altering the properties of nonconductors and, as a matter of fact, displacement damage in ionic compounds and ceramics may be larger than in most metals.[8] However, the mechanical properties of such substances, during or after irradiation have not yet been explored.

An examination of Table I shows that the elastic constants are the only mechanical characteristics which are well understood. From a fundamental standpoint it is naturally desirable to study the effect of radiation on properties which are best understood. In general, a bulk property is not expected to be changed appreciably by the displacement of a relatively few number of atoms. Recent theoretical work by Dienes[49] indicates, however, that the presence of interstitial atoms will result in appreciable increase in the elastic moduli, at least for close packed metallic lattices. In such substances the elastic constants are determined primarily by the repulsive interactions of the closed ion shells. This potential is of an exponential nature and varies extremely rapidly with interatomic distance. As the interaction distance is shortened, for example, by creating an interstitial, the energy of the system increases sharply on the repulsive side of the potential curve. The creation of vacancies by removal of atoms results essentially in the destruction of some normal interactions. Thus, one expects the influence of the interstitials to outweigh heavily the effect of the vacancies.

Detailed calculations confirmed the above picture and led to the following conclusions. The presence of a small fraction of interstitials and vacancies results in large increases in the elastic moduli of copper, of the order of 5–7 percent for one percent interstitial. Lattice vacancies alone were found to decrease the elastic moduli, the change being essentially a bulk effect. Consequently, increases in the elastic moduli are to be attributed primarily to the presence of interstitial atoms. The theoretically predicted effects should be easily observable experimentally in copper and similar metals provided thermal annealing is prevented. Preliminary experiments by Bowman and Tarpinian[50] of North American Aviation indicate that Young's modulus of copper increases by as much as 10 percent upon heavy cyclotron irradiation.

This type of an investigation is of interest in yielding information about one of the fundamental mechanical properties. Further, the results strongly suggest that changes in elastic constants may serve as a useful and powerful tool for distinguishing between interstitial atoms and vacant lattice sites. This latter point is of some importance since these two types of disturbances are very difficult, if not impossible, to distinguish by electrical or thermal resistivity measurements (which change drastically upon irradiation) since they are both scattering centers for electron and thermal waves. Much more work is required for proper evaluation of the suggestions.

In contrast to bulk properties, some structure sensitive mechanical properties are very drastically changed by irradiation. The interpretation of these phenomena is much more intricate and difficult for two reasons. These properties are not well understood even in the absence of radiation (Table I), and the phenomena themselves are probably quite complex.

No studies have been published concerning the influence of nuclear radiations on anelastic properties.

[47] E. Saibel, Trans. Am. Inst. Mining Met. Engrs. 171, 639 (1947).
[48] M. Born, J. Chem. Phys. 7, 591 (1939).
[49] G. J. Dienes, Phys. Rev. 86, 228 (1952); 87, 666 (1952); see also N. R. F. Nabarro, Phys. Rev. 87, 665 (1952).
[50] F. Bowman and M. Tarpinian (private communication, February, 1952).

Many interesting and revealing experiments could be done utilizing our present knowledge of these properties and some will be suggested later on.

A considerable amount of work has been done on the influence of radiation on the plastic properties of metals. It is well established by now that increase in hardness resulting from irradiation is a very general phenomenon.[51,52] The work of Blewitt, Coltman, and Sherrill[53] on single crystals of copper shows very clearly that the major effect of radiation is to alter greatly the critical shear stress. Stress-strain curves obtained on irradiated and unirradiated crystals are shown in Fig. 1. The critical shear increased from 0.241 to 2.0 kg/mm² for a fast neutron flux of about 2×10^{18} nvt.‖ After appreciable plastic deformation the differences in the stress-strain curve essentially disappear. The discontinuities at a shear strain of about 0.6 are due to an additional irradiation of about 2×10^{18} nvt. It is clear that after cold work the effect of irradiation is much smaller than for annealed samples. It is also clear from Fig. 1 that there is no one-to-one correspondence between cold work and radiation hardening.

The shape of the stress-strain curve after irradiation is similar to that produced by solid-solution hardening, for example, by zinc in copper (α-brass).[54] Further, critical shear stress *vs* bombardment curves, shown in Fig. 2, indicate strong similarity to shear hardening by dissolved impurities.[55] Jamison and Blewitt[56] have also investigated the slip lines formed upon cold-working irradiated and unirradiated single crystals of copper and observed pronounced differences in the two cases. The slip characteristics of the irradiated crystals were similar to those of α-brass.

These experiments indicate a very strong similarity between impurity and radiation hardening. Current theories of impurity hardening[57–60] are based on the interaction between dislocations and impurity atoms. The stress field set up by an impurity atom which does not fit perfectly into the lattice interacts with the stress field of a dislocation in such a way as to impede the motion of the dislocation resulting in hardening of the material. Vacant lattice sites and interstitial atoms are expected to behave in a similar way. The simplest form of the theory indicates that the hardening is proportional to the square of the "misfit"[57] of the lattice disturbance. Interstitial atoms introduce a much larger disturbance in the lattice than vacant lattice sites and one expects, therefore, that the former will be far more effective in raising the critical shear stress.

Theory and experiment are apparently in quite good qualitative agreement. A quantitative description, however, is much more difficult. Comparison of radiation hardening to those impurities which are well described by the theory (for example, α-brass) indicates that unreasonably large distortions would have to be ascribed to the interstitial atoms to explain the results. Thus, one may have to postulate that vacancies and interstitials are present in the form of clusters, probably in the neighborhood of dislocations, and contribute to the hardening by means of a Cottrell[59] type of anchoring of the dislocations. On the other hand, there are alloy systems, such as Cd in Zn, in which the critical shear stress is much higher than predicted by theory. Similarly, critical shear stress of the alkali halides is extremely sensitive to small amounts of impurities.[61] Thus, the possibility that isolated interstitials and vacancies may account for radiation hardening cannot be excluded at the present time. Further detailed experimental and theoretical work will be required for clarification of the above points.

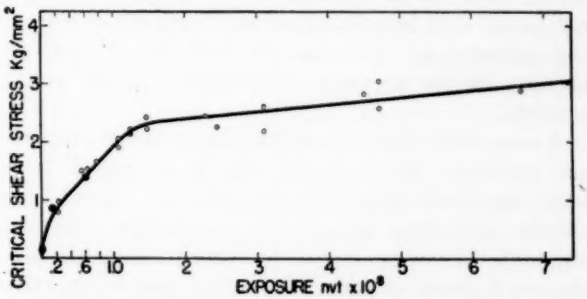


FIG. 2. Effect of neutron irradiation on the critical shear stress of copper (from reference 55).

Among the time-dependent plastic phenomena creep has received some attention. It is very difficult to estimate theoretically the effect of radiation on creep rate.[7,62] One can argue that the imperfections introduced in the lattice should impede the motion of dislocations and thereby reduce the creep rate. On the other hand, creep may be more rapid during irradiation since enough of the atoms may be kept stirred up by the irradiation to allow the stress to relax more rapidly. Andrade[63] measured the creep rate of cadmium under α-particle bombardment from polonium and found an

[51] D. S. Billington and S. Siegel, Metal Progr. **58**, 847 (1950).

[52] Martin, Austerman, Eggleston, McGee, and Tarpinian, Phys. Rev. **81**, 664A (1951); see also unclassified portions of AEC Report NAA-SR-75 (1951).

[53] Blewitt, Coltman, and Sherrill, Unclassified AEC Report TID-5015; T. H. Blewitt and R. R. Coltman, Phys. Rev. **82**, 769A (1951).

‖ nvt is used to designate the number of neutrons per cm² incident on the sample. nv is the flux in neutrons per cm² per second and t is the irradiation time.

[54] See E. Schmid and W. Boas, *Plasticity of Crystals* (F. A. Hughes and Company, London, 1950), pp. 133–145.

[55] T. H. Blewitt (private communication).

[56] R. E. Jamison and T. H. Blewitt, Phys. Rev. **86**, 641A (1952).

[57] N. F. Mott and F. R. N. Nabarro, *Bristol Conference on Strength of Solids* (The Physical Society, London, 1948), pp. 1–20.

[58] F. R. N. Nabarro, Proc. Phys. Soc. (London) **58**, 669 (1946); **59**, 256 (1947); Proc. Roy. Soc. (London) **175A**, 519 (1940).

[59] A. H. Cottrell, Pittsburgh Symposium on Plastic Deformation of Crystalline Solids (U. S. Office of Naval Research, 1950), pp. 60–77.

[60] N. F. Mott, "Mechanical strength and creep in metals," in *Imperfections in Nearly Perfect Crystals*, Editor, W. Shockley (John Wiley and Sons, Inc., New York, 1952), pp. 173–197.

[61] See reference 54, pp. 236–237.

[62] F. R. N. Nabarro, *Bristol Conference on Strength of Solids* (The Physical Society, London, 1948), pp. 75–91.

[63] E. N. da C. Andrade, Nature **156**, 113 (1945).

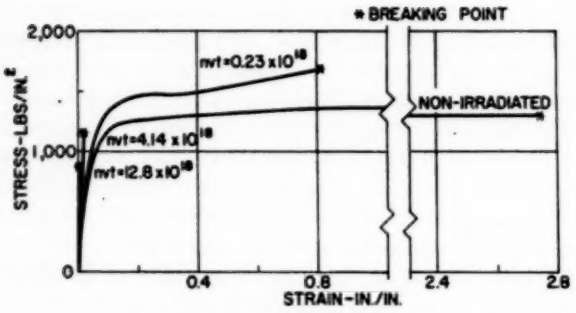


FIG. 3. Stress-strain curves for pile irradiated polyethylene (from reference 72).

appreciable increase in the creep rate, particularly during the early stage of the creep process. More recently Witzig[64] found no change in the creep rate of copper while the sample was bombarded by deuterons. Thus, experiments so far are not consistent. Andrade's results may be explicable in terms of surface effects. It is known that the creep rate of some metals is extremely sensitive to surface conditions.[65] Witzig used rather low rates of irradiation and his null results may not be valid at heavier exposure. Creep experiments in the presence of radiation are difficult to carry out because of the accurate temperature control that is required.

A somewhat different aspect of mechanical properties and radiation effects should also be discussed. It has been noted that there are similarities between cold-work (strain hardening) hardening and radiation hardening, although there is no one-to-one correspondence between the two types of damage. This sort of parallelism carries through in other properties also. Electrical resistivity, for example, increases as a result of irradiation as well as of cold work. It is natural to suggest an analogy between cold work and radiation damage based on the notion that the properties of crystals are a unique function of the atomic positions no matter how the atoms got there.[7,8] In many cases this has been a fruitful notion in guiding experimentation. However, the analogy must break down if the nature of the disturbances introduced by cold work and by radiation is quite different. First of all, plastic deformation involves the motion and multiplication of dislocations, while irradiation is not expected to introduce such an organized lattice disturbance. Some significant differences in behavior are to be expected on this basis. X-ray line broadening, for example, which is appreciable in metals cold-worked at room temperature and is successfully interpreted on the basis of internal stresses,[66] has not yet been observed in irradiated metals. Whether the negative results obtained to date are attributable to differences in the types of disturbances introduced or primarily to fast annealing of the radiation effects cannot yet be determined.[67] Further, differences are also expected in the vacancy—interstitial type of disturbances left behind in the lattice. Recent theoretical work by Seitz[33] indicates that vacancies are produced by cold work. Interstitials may also be introduced by plastic deformation, but it requires much higher energies to produce interstitials than vacancies, at least in metals,[68,69] so that one may expect the number of interstitials created by cold work to be rather small. No such restrictions apply, of course, to the production of interstitials by fast particle irradiation, and one may expect corresponding differences in the behavior of cold-worked and irradiated materials. Indeed, significant differences in annealing behavior have been observed which may be explicable by such considerations. Bowen and co-workers[70] have done the most detailed work of this kind but much more needs to be done before full interpretation is possible. It is quite certain, however, that one can learn a great deal about radiation effects and the nature of plastic deformation from careful comparative studies of these two types of crystalline damage.

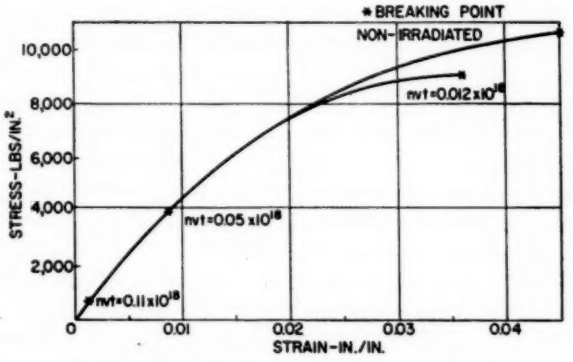


FIG. 4. Stress-strain curves for pile irradiated Lucite (from reference 72).

A complex set of phenomena may arise when alloys are irradiated. Either via the production of vacancies and interstitials or via thermal spikes, local atomic interchanges (microscopic diffusion) may be promoted by irradiation. As far as the mechanical properties are concerned the precipitation hardening alloys are of particular interest. The copper-beryllium system has been studied by Billington and Siegel,[5] Martin and co-workers,[52] and Murray and Taylor.[71] Radiation hardening in solution annealed samples of this alloy is much greater than in pure copper. Resistivity increases

[64] W. F. Witzig, Unclassified AEC Report, WAPD-43 (1951).
[65] See, for example, D. J. Phillips and N. Thompson, Proc. Phys. Soc. (London) **63B**, 839 (1950).
[66] B. L. Averbach and B. E. Warren, J. Appl. Phys. **20**, 1066 (1949).
[67] B. E. Warren, Unclassified AEC report, NYO-767, (1951).
[68] H. B. Huntington and F. Seitz, Phys. Rev. **61**, 315 (1942); **76**, 1728 (1949).
[69] H. B. Huntington, Phys. Rev. **61**, 325 (1942).
[70] Bowen, Eggleston, and Kropschot, J. Appl. Phys. **23**, 630 (1952); see also paper given to the Society of Rheology at the Chicago Anniversary Meeting of AIP, October, 1951 (unpublished).
[71] G. T. Murray and W. E. Taylor, Unclassified AEC report TID-5017 (1951).

are also quite large. In general, the more beryllium there is in solution the greater is the effect of irradiation. These findings suggest that segregation or precipitation nucleation processes occur during irradiation. X-ray evidence is not yet consistent or conclusive and reliable interpretation of these experiments is not yet possible.

3. Radiation Effects in Molecular Solids—Mostly High Polymers

From the general discussion given in Section II one anticipates that the primary effects of nuclear radiations in molecular solids will be the result of ionization and/or excitation and will be essentially the same for all types of high energy radiation. Ions and free radicals are expected to be produced which then bring about chemical reactions (bond breakage and formation) thereby altering the properties of the substance. As stated before, high polymers are the only molecular solids whose mechanical properties have been investigated to any extent with respect to radiation effects.

The investigation of radiation effects in high polymers was begun primarily for practical reasons. The stability

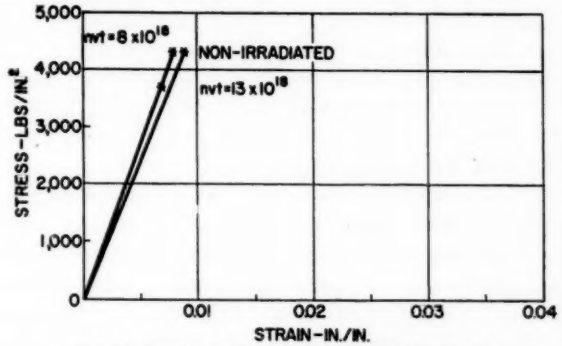


FIG. 5. Stress-strain curves for pile irradiated polystyrene (from reference 72).

of these substances is of considerable importance in reactor technology. At the present time most of the available information is of an empirical nature, which has been summarized recently by Sisman and Bopp.[72] In general, discoloration, brittleness, chemical decomposition, and many other changes may result from pile exposure of plastics. Sisman and Bopp have been able to classify empirically a large number of plastic materials into 12 groups in order of decreasing radiation resistance. Mineral filled phenolics apparently withstand radiation best. Among the unfilled polymers polystyrene is outstanding. In both cases the well-known stabilizing influence of benzene rings is apparent. At the other end of the scale methyl methacrylate, halogenated polyethylenes and cellulosics are extremely poor in radiation resistance.

As far as the mechanical properties are concerned, changes in stress-strain curves as a function of irradiation are the most revealing among the properties studied by Sisman and Bopp (stress-strain, impact and

[72] O. Sisman and C. D. Bopp, Unclassified AEC report, ORNL-928(1951).

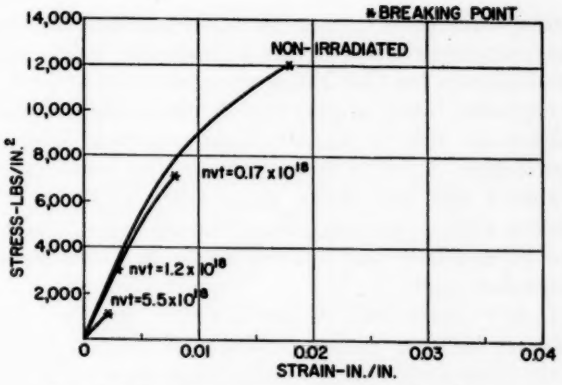


FIG. 6. Stress-strain curve for pile irradiated paper base phenolic (from reference 72).

shear strength, hardness). One generalization may be made which is almost universally valid for all the high polymers they studied. The elongation at break is drastically reduced by irradiation, i.e., the material loses its plasticity and can no longer be cold-drawn. This change is strikingly illustrated by the curves of Fig. 3 for polyethylene. The irradiation dosages are indicated along the curves in terms of the total number of neutrons incident of the sample per cm² (nvt). Extensibility is rapidly lost even at quite low dosages. Tensile strength is not changed significantly, but the stress-strain curve is drastically altered as the radiation is increased. The material finally becomes extremely brittle with practically no extensibility. Sisman and Bopp state that polyethylene and nylon become so brittle that the corners of the specimens chip off.

The response of methylmethacrylate to irradiation is of interest. The resistance of this material is poor as shown in Fig. 4. It is noteworthy, however, that the shape of the stress-strain curve is hardly changed. Irradiation alters the state of the material in such a way that failure occurs at decreased stresses and elongations with the points of failure falling along the same curve.

Much higher resistance to irradiation is shown by polystyrene, Fig. 5. Only minor changes in elongation and stress are observed at the exposures employed in these experiments. Mineral filled phenolics behave in a very similar way. However, if an organic filler is used, fast deterioration upon irradiation is observed. In Fig. 6 stress-strain data are given for a paper base phenolic. Drastic reduction in stress and elongation at failure is observed with relatively minor changes in the shapes of the stress-strain curves.

It is evident that understanding of these phenomena will come from a careful correlation of changes in physical properties with alterations in structure. Studies of a more fundamental nature have been started by Davidson and Geib[73] on rubbers and by Dole[74] on polyethylene. On the basis of plasticity and

[73] W. L. Davidson and I. G. Geib, J. Appl. Phys. **19**, 427 (1948).

[74] M. Dole, "Effect of radiation on colloidal and high polymeric substances," Army Chemical Center Symposium on Chemistry and

stress-strain measurements Davidson and Geib conclude that natural rubber increases in molecular weight upon irradiation in the Oak Ridge pile while polyisobutylene is degraded. They suggest that free radicals in natural rubber are able to initiate chain reactions involving the double bonds and apparently the cross-linking achieved this way more than compensates for the broken chains. In polyisobutylene no chain reaction can be effected, and the end result is a decrease in molecular weight.

Dole[74] chose polyethylene as the simplest high polymer for investigating radiation effects. Irradiations in the presence of both air and vacuum were carried out in the Argonne heavy water pile. The polyethylene used was in the form of sheet film. As far as the changes in mechanical properties are concerned Dole confirmed the results of Sisman and Bopp. Chemical and infrared absorption analyses showed that considerable unsaturation occurred in the vacuum irradiated samples. Oxidation was the major effect in the material irradiated in air. Cross-linking may have also occurred. Molecular weights and molecular distributions have not yet been measured nor have the electrical properties been investigated. Tentatively Dole suggests that the major chemical changes occur upon neutralization of the positive charge which is situated somewhere along the chain as a result of electron depletion by the irradiation. Free radicals are produced as a result of the split in the C−C bond leading to unsaturation and to oxidation in the presence of air.

IV. SUGGESTED EXPERIMENTS

On the basis of the discussion given in previous sections the writer would like to suggest several experiments which should further our knowledge of the fundamental problems involved.

It should be pointed out that from these experiments one will learn not only about radiation effects but also about the fundamental solid-state physics problem of relating physical properties of materials to lattice imperfections and structural changes.

For crystalline materials the following experiments appear to be fruitful.

(a) A careful study of the elastic constants of irradiated crystals should be made. Preliminary experiments indicate that elastic constants increase upon irradiation in accordance with the theory.[50] It appears that quite heavy irradiations, preferably performed at low temperature, are required for sufficient sensitivity. The experiments should be done on single crystals for simplicity of interpretation, particularly with respect to annealing. The major usefulness of this experiment is to indicate the presence of interstitial atoms and to distinguish them from vacant lattice sites.

(b) It is known from Nowick's[22] work that vacancies may be frozen in and then detected by internal friction measurements in suitable alloy systems. It would be of great interest to produce vacancies by means of fast particle bombardment and measure their concentration by internal friction. The annealing of the lattice disturbances can also be followed by this technique. The major requirement of this experiment is that the two components of the substitutional alloy differ sufficiently in size to lead to anelastic effects.

(c) Critical shear stress of irradiated alkali halide crystals should be determined as a function of irradiation and temperature. The critical shear stress of these crystals appears to be extremely sensitive to the presence of very small amounts of impurities and one expects interstitials and vacancies to act in a similar manner.

(d) Detailed annealing studies of critical shear stress hardening by radiation should lead to an accurate determination of the mobilities and activation energies of the disturbances responsible for radiation hardening. From such data it should be possible to deduce the size of the disturbances and hence settle the question of whether isolated disturbances or clusters of disturbances are present in the lattice.

For high polymers the following experiments are suggested:

(a) The characteristics of the stress-strain curve before and after irradiation should be carefully related to structural changes such as double bond formation and oxidation. This is probably best done by means of infrared analytical techniques.

(b) The altered mechanical properties should be correlated with possible changes in molecular weight and molecular weight distribution of the substance.

(c) Both in rubbers and elastomers creep and stress relaxation characteristics may often be interpreted on the basis of bond breakage and formation.[16,75–77] It would be of great interest to investigate these time-dependent properties in the presence of nuclear irradiations. One anticipates that quite large effects may be produced. In many ways the polymer may be altered in a manner by irradiation which is under better experimental control than, say, during high temperature degradation.

(d) It may well be profitable to compare critically the discoloration almost invariably produced by irradiation to that effected by thermal aging.

(e) Many of the suggestions given above are applicable to inorganic glasses with minor modifications. Apart from some work on color changes[78] this class of solids has received very little attention as far as irradiation phenomena are concerned.

Note added in proof:—Attention is called to two recent papers by Charlesby (Proc. Roy. Soc. (London) **215A**, 187 (1952); Nature **171**, 167 (1953), published after this paper was submitted, in which detailed experiments, pointing to cross-linking in polyethylene by pile radiation, are discussed.

Physics of Radiation Dosimetry, NP-3237 (1950, unclassified), pp. 120–148.

[75] A. V. Tobolsky and R. D. Andrews, Jr., J. Chem Phys. **13**, 3 (1945).

[76] M. D. Stern and A. V. Tobolsky, J. Chem. Phys. **14**, 93 (1946).

[77] Andrews, Hofman-Bang, and Tobolsky, J. Polymer Sci. **3**, 669 (1948).

[78] G. J. Monk, Unclassified AEC report, ANL-4536 (1950).

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

A Test of the Theory of Secondary Viscoelastic Stress

M. MOONEY
General Laboratories, United States Rubber Company, Passaic, New Jersey
(Received February 16, 1953)

Experimental data published by Garner, Nissan, and Wood on secondary stresses in a viscoelastic liquid have been analyzed in terms of the superelasticity theory of secondary stresses as developed by Mooney. The theory was slightly modified to be consistent with the reported viscosity rate of shear behavior of the test liquid. With this modification the theory agreed within experimental error with observed normal stresses at various radii on a stationary circular plate mounted parallel to a rotating plate.

One reported qualitative experimental result is in disagreement with the theory, but this particular result is questioned. Other theories of secondary stress are briefly discussed.

INTRODUCTION

IN a recent paper by Garner, Nissan, and Wood[1] some quantitative experimental results are reported on the pressures developed in a viscoelastic liquid subjected to continuous shear between a pair of parallel circular plates in relative rotation. The experimental arrangement, similar to one first described by K. Weissenberg,[2] consisted of a flat-bottomed beaker driven in continuous rotation about its axis and a centrally mounted stationary circular plate a short distance above the bottom of the beaker. The normal stress of the sheared liquid on the stationary plate was measured by means of a series of manometer tubes connecting with small port holes at various radial distances on the plate. The test liquid for the data reported consisted of a 7 percent solution of aluminum stearate in petrol, peptized by mixed xylenols.

Shortly after the appearance of this paper by Garner, Nissan, and Wood (hereafter to be referred to as GNW) the present writer published a theoretical paper[3] (hereafter to be referred to as M) dealing with secondary viscoelastic stresses, in general, and covering the GNW experimental arrangement as one special case. At the time this theoretical paper was in preparation the author was, unfortunately, unaware of the GNW experimental work. The present paper is therefore now written to test the theory with the experimental data.

REVIEW AND MODIFICATION OF THE THEORY

The author's theory is based on the established general stress-strain[4] relations of a superelastic material in finite deformation. The connection with superelasticity comes from the assumption that in a viscoelastic liquid subjected to continuous simple shear the elastic structure in the liquid is in a state of constant shear stress τ_s, the magnitude of τ_s being determined by the balance between the rate of stress development by the imposed continuous shear and the rate of stress relaxation characteristic of the material in the given state of stress. In this constant state of shear stress there is a corresponding constant shear strain s, which theoretically is the elastic recovery that would be observed if the shear stress were instantaneously removed and the recovery could take place instantaneously or too fast for further relaxation. If G is the modulus of rigidity of the elastic structure, the relationship is

$$\tau_s = Gs. \tag{1}$$

It is the magnitude of s which determines the ratio of the primary shearing stress τ_s to the secondary, normal deviatory stresses which develop on the three orthogonal planes corresponding to the orientation of the shear axes. If the simple shear is such that, in relation to reference axes x, y, z, the xz planes are moving in the positive x direction with velocity gy, then the stress relations are[3]

$$\begin{cases} \sigma_x = -p + \left(\dfrac{G}{2} + \dfrac{H}{6}\right)s^2, \\ \sigma_y = -p - \left(\dfrac{G}{2} - \dfrac{H}{6}\right)s^2, \\ \sigma_z = -p - \dfrac{H}{3}s^2, \end{cases} \tag{2}$$

$$\tau_{xy} = Gs. \tag{3}$$

Here the σ_i are the normal stresses, p the mean pressure, τ_{xy} the shearing stress, and H is the second superelasticity modulus as defined by Mooney.[4] The deviatory stresses are $\sigma_i + p$.

In order to apply the above equations to experimental data, it is necessary to have a relationship between g, the rate of continuous shear, and s, the elastic shear, and relationships also between g and the elastic moduli G and H if the moduli are not constant. The simplest postulates would be that s is proportional to g, while G and H are constants. These postulates imply Newtonian viscosity for the liquid, but GNW report for

[1] Garner, Nissan, and Wood, Trans. Roy. Soc. (London) **243**, 37 (1950).
[2] K. Weissenberg, Nature **159**, 310 (1947).
[3] M. Mooney, J. Colloid Sci. **6**, 96 (1951).
[4] M. Mooney, J. Appl. Phys. **11**, 582 (1940); also L. R. G. Treloar, *The Physics of Rubber Elasticity* (Oxford University Press, London, 1949).

their material a viscosity relationship approximately of the power form

$$\eta = B_g^{-c}, \tag{4}$$

where $c \sim 0.6$ or 0.7. A non-Newtonian case was analyzed in M, but the assumed $\eta - g$ relationship was entirely different.

For present purposes let us therefore assume the following relationships for the steady state of continuous shear at rate g:

$$s = g/\rho, \quad \rho = \rho_1 g^p, \quad G = G_1/g^q, \quad H = H_1/g^q, \tag{5}$$

$$\tau = g\eta_1 + \tau_e, \quad = g\eta_1 + Gs. \tag{6}$$

Here ρ is the relaxation rate, and ρ_1, G_1, and H_1 are constants. τ is the total shear stress exerted by the viscoelastic solution, $g\eta_1$ being that which is due to the solvent, and τ_e that which is due to the elastic gel structure. η_1 is proportional to the viscosity of the solvent but is presumably also a function of the soap concentration. The observable viscosity is

$$\eta = \frac{\tau}{g} = \eta_1 + \frac{Gs}{g} = \eta_1 + \frac{G_1}{\rho_1} g^{-(p+q)}. \tag{7}$$

From this equation it is apparent that the viscosity follows Eq. (4) approximately if η_1 is small in comparison with $G_1/\rho_1 g^{p+q}$.

Consider now the experimental arrangement of GNW, for which cylindrical coordinates r, θ, z are appropriate, with z being measured downward from the upper, stationary plate. Corresponding to Eqs. (2) and (3) we then have

$$\begin{cases} \sigma_r = -p - \dfrac{H}{3} s^2, \\ \sigma_\theta = -p + \left(\dfrac{G}{2} + \dfrac{H}{6}\right) s^2, \\ \sigma_z = -p - \left(\dfrac{G}{2} - \dfrac{H}{6}\right) s^2; \end{cases} \tag{8}$$

$$\tau_{\theta z} = -Gs. \tag{9}$$

As shown in M, the radial normal stress then obeys the differential equation

$$\frac{\partial \sigma_r}{\partial r} = \frac{G+H}{2r} s^2 - D\omega^2 r, \tag{10}$$

where D is the density and ω is the local angular velocity of the liquid. We have also

$$\begin{cases} g = r\dfrac{\Omega}{Z}, \\ \omega = \dfrac{z}{Z}\Omega, \end{cases} \tag{11}$$

where Ω is the angular velocity of the rotating plate and Z is the plate separation.

With the substitutions indicated by Eqs. (5) and (11), Eq. (10) becomes

$$\frac{\partial \sigma_r}{\partial r} = \frac{G_1 + H_1}{2\rho_1^2 r} \left(\frac{r\Omega}{Z}\right)^{2-2p-q} - D\frac{\Omega^2 z^2}{Z^2} r. \tag{12}$$

Integration from 0 to r gives

$$\sigma_r(r, z) - \sigma_r(0, z) = \frac{G_1 + H_1}{2\rho_1^2} \left(\frac{r\Omega}{Z}\right)^{2-2p-q} \times \frac{1}{2-2p-q} - \frac{D\Omega^2 z^2 r^2}{2Z^2}. \tag{13}$$

What is observed in the experiment is not $\sigma_r(r, z)$ but $\sigma_z(r, 0)$ as indicated by the liquid levels in the manometer tubes. From Eqs. (8.1), (8.3), (5), and (11.1) it is found that

$$\left.\begin{aligned} \sigma_z(r, z) - \sigma_r(r, z) &= -(G-H)s^2/2, \\ &= -\frac{G_1 - H_1}{2\rho_1^2} \left(\frac{r\Omega}{Z}\right)^{2-2p-q}. \end{aligned}\right\} \tag{14}$$

The argument z may be introduced in the left member of this equation because the right member is independent of z and the equation is valid for any z. It is clear also that

$$\sigma_z(0, z) - \sigma_r(0, z) = 0. \tag{15}$$

This means that the stress system at the axis is a uniform, hydrostatic pressure. From Eqs. (13), (14), and (15) if follows that

$$\sigma_z(r, z) - \sigma_z(0, z) = \left[-\frac{G_1 - H_1}{2\rho_1^2} + \frac{G_1 + H_1}{2\rho_1^2(2-2p-q)}\right] \times \left(\frac{r\Omega}{Z}\right)^{2-2p-q} - \frac{D\Omega^2 z^2 r^2}{Z^2}. \tag{16}$$

At the stationary plate, where the manometers are attached, $z = 0$. Also, the normal pressure P is the negative of the normal stress σ_z. Hence,

$$P(0) - P(r) = \left[\frac{G_2 + H_1}{2\rho_1^2(2-2p-q)} - \frac{G_1 - H_1}{2\rho_1^2}\right] \left(\frac{r\Omega}{Z}\right)^{2-2p-q}. \tag{17}$$

COMPARISON WITH EXPERIMENT

This equation predicts that the observed pressure difference should vary as a power of the shear rate g, or $r\Omega/Z$. Accordingly, these pressure differences in cm of the solution were computed from the data of GNW, Table 8, and are shown in Fig. 1, plotted on logarithmic coordinates against $r\Omega/Z$. Each series of points con-

nected by a broken line corresponds to a series of r values for a particular pair of Ω- and Z-values.

The data are scattered, thus reflecting the experimental difficulties in this field of research; but the scattering seems to be quite random. Hence we may conclude that, within experimental error, there is agreement between theory and experiment.

The heavy line in Fig. 1, drawn to represent all the points, has a slope 0.775. Consequently,

$$2p+q=1.225. \tag{18}$$

If we neglect η_1 in Eq. (7), then Eq. (4) indicates that

$$p+q=0.65\pm0.05. \tag{19}$$

These last two equations indicate that within the precision of this procedure

$$q=0. \tag{20}$$

This conclusion can only be accepted as tentative at present. It is *a priori* questionable because it indicates no thixotropy and no breakdown of structure by shearing. Furthermore, in the secondary stress measurements a steady state was approached very slowly, and measurements were made only after 2 hours continuous running at constant speed, whereas in the viscosity measurements leading to Eq. (4), the rolling ball method was used[5] and each measurement involved only a very limited deformation of the test material. There is, therefore, some doubt whether the value given for c in Eq. (4) is the correct value to be used in the present calculations.

It would seem desirable in dealing with these viscoelastic solutions that all physical tests should be made under similar conditions and similar immediate past history of the sample. Furthermore, from the theorist's point of view, the experimental methods should be elaborated to give enough independent measurements to determine all the parameters entering into the theory. Thus, besides primary (viscosity) and secondary stress, birefringence and angle of extinction could be measured. The last measurement would permit a direct calculation of the steady shear s. Birefringence, according to Trelor's work,[6] would presumably be proportional to the difference of the two extreme principal elastic stresses. Modulus and relaxation rate could be measured in the state of continuous shear if an oscillation were superimposed on the continuous shear. Just to mention these various possibilities shows that this new experimental field of secondary stresses has a number of interesting avenues of exploration which eventually should lead to a check of all details of the theory so far developed.

A few words should be added concerning the failure to find pressure differences in two of the GNW experimental arrangements, those shown by the their Figs. 12(a) and 12(c). In Fig. 12(a) a hollow cylinder containing the soap solution rotates continuously about a slightly smaller stationary solid cylinder, and a manometer in the stationary cylinder connects to a port in the cylindrical surface. It was found that the liquid level in the manometer showed no reproducible measurable difference from the level in the containing cylinder.

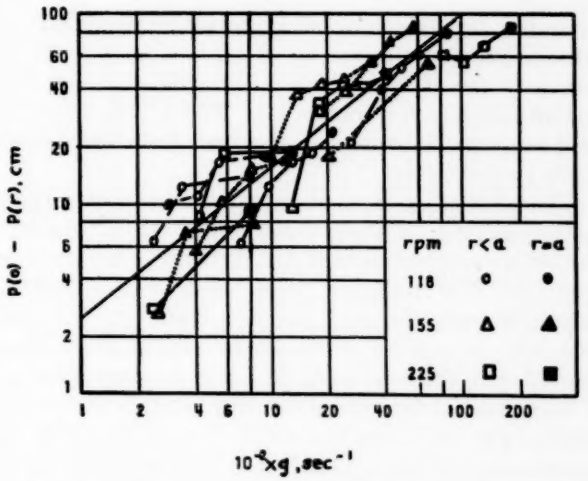


FIG. 1. Pressure difference *versus* $r\Omega/h$, or g, the local rate of shear. The value of a, required for plotting the last point in each series, is given as 3.24 cm by A. H. Nissan in a personal communication. The long solid line corresponds to Eq. (17) with $2-2p-q=0.775$.

This negative result is explainable by the present theory. From Eqs. (8.1) and (8.3) the normal stress difference is

$$\sigma_r-\sigma_z=(G-H)s^2/2. \tag{21}$$

This equation may be used as a valid approximation for the experiment considered, since the difference between the cylinder radii is small compared with the mean radius. (The appropriate exact equation would include a term proportional to $gG(r_2-r_1)/(r_2+r_1)$ the r_i being the cylinder radii.) The approximate Eq. (21) shows that no pressure difference is to be expected if

$$H=G. \tag{22}$$

This last relationship can be readily accepted as possible for it is predicted by the kinetic theory of elasticity for a cured elastomer, as developed by Wall,[7] and is found by Copeland and Mooney[8] to be approximately correct for several elastomers.

The other GNW experimental arrangement, that of their Fig. 12(c), consisted of a plate and a cone in relative rotation, the cone apex lying in the plane of the plate. This arrangement has already been analyzed in M with the conclusion that there should be observed a pressure difference between any two radial distances,

[5] Wood, Nissan, and Garner, J. Inst. Petroleum **33**, 71 (1947).
[6] L. R. G. Treloar, Trans. Faraday Soc. **43**, 277 (1947).
[7] F. T. Wall, J. Chem. Phys. **10**, 485 (1942); Rubber Chem. Technol. **15**, 806 (1942).
[8] L. E. Copeland and M. Mooney, J. Appl. Phys. **19**, 450 (1948).

r_1 and r_2, the difference being proportional to $\log r_2/r_1$. Modification of the theory to harmonize it with the viscosity power law of Eq. (4) would not alter this logarithmic pressure-radius law. The GNW failure to find pressure differences in this case, therefore, is contrary to the theory.

The present writer has repeated these two experiments, using a 10 percent solution of GR-S (standard butadiene-styrene copolymer) in benzene as the test material. The negative result with the cylindrical apparatus was checked, but not that with the cone and plate apparatus. In the latter case the plate speed was 43 rpm, the cone angle 75°, and the cone diameter 2 in. A manometer connected with a small port at the apex of the cone. The liquid rose the full height of the manometer, 12 in., and continued to flow slowly out at the top. From this positive result it is clear that the GNW negative result with their aluminum soap solution in petrol cannot be taken as typical of all viscoelastic materials.

GNW report experimental results obtained by several other experimental methods, but these other methods are not easily analyzed in terms of the present theory. With regard to the results given by these methods, it can only be said that there is no apparent conflict with the theory.

OTHER THEORIES

This superelasticity theory of secondary stress could be classed as a special case under a very general rheological theory developed by Oldroyd;[9] but Oldroyd himself does not give any equations directly applicable to secondary stress measurements.

A theory presented in the GNW paper is based on considerations of the free energy of the soap solution when the colloidal structure is broken up into individual soap and peptizer molecules; but this theory must be rejected, in the present author's opinion, at least in its present form. Several of the GNW detailed postulates could be questioned or criticized; but the most convincing argument is based on their final, theoretical pressure-radius equation, which has the form

$$P(r)=P(0)-K\left(\frac{\Omega}{r}\right)^{1.4}/r^{1.6}.$$

According to this equation, $P(r)$ should increase with r, which is contrary to published experimental data including the GNW data. Also, the equation has a discontinuity at $r=0$, whereas the experimental data seem to be regular at this point. It is undoubtedly true, as assumed by GNW, that a viscoelastic liquid has free energy when in a state of continuous shear; but it does not necessarily follow that the free energy is due to an increase in the concentration of fragments of the broken colloidal structure. There would also be free energy in the elastically strained colloidal structure that remains unbroken. According to the above analysis, it is the latter, the mechanical free energy, that is important in the present case, not the chemical or concentration free energy.

Another theory in this field has been developed by R. S. Rivlin,[9] who uses the term "normal stress" for what has here been called "secondary stress." The Rivlin theory postulates an incompressible viscoinelastic liquid that is isotropic in a state of rest. For such a material the most general possible relations between the stress and strain-velocity components are developed. The resulting formula applicable to the relatively rotating parallel plates, Rivlin equation (15.15), contains the normal stress coefficient ψ. This coefficient may be any function of the scalar invariants of the deformation rates; and if we assume the appropriate power form for this function, the equation for the z-stress difference between radii 0 and r can be put in the form

$$P(0)-P(r)=C(r\Omega/Z)^n. \tag{23}$$

This is similar in form to Eq. (17) above, and hence it is seen that the data in Fig. 1 are explainable by the Rivlin theory also. On the other hand, let us consider now the rotating cylinders experiment. Rivlin's equation (16.7), in present notation, is

$$\sigma_r-\sigma_z=\psi s^2/2. \tag{24}$$

Since ψ, the normal stress coefficient, cannot vanish if the data of Fig. 1 are to be explained, Eq. (24) stands in disagreement with Eq. (21), when $H=G$, and stands in disagreement also with the experimental results obtained with the rotating cylinders by both GNW and the present author. Both of the theories under discussion, incidentally, disagree with the GNW results with the cone and plane experiment; and hence on this basis there is no choice between the theories. They are both wrong if the GNW results with the cone and plane are valid. In view of the present writer's results with the same test, the facts seem to be in doubt and more experimental work is clearly desirable.

Note added in proof:—In a recent article by H. W. Greensmith and R. S. Rivlin, Phil. Trans. **245A 899**, 399 (1953), some further experimental results are reported with the arrangement shown in GNW Fig. 12(a) and discussed above in connection with Eq. (21). The Greensmith and Rivlin results are in agreement with GNW and show again that $\sigma_r-\sigma_z=0$.

This conclusion from the experimental results is overlooked by Greensmith and Rivlin; but to the present author it seems inescapable. A further conclusion is that the Rivlin (see reference 9) theory of normal stress in a viscoinelastic liquid is not general enough to explain completely the behavior of a viscoelastic liquid.

[9] R. S. Rivlin, Proc. Roy. Soc. (London) **A193**, 260 (1948). See also, I. Braun and M. Reiner, Quart. J. Mech. and Appl. Math. 5 (1952).

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Dynamic Mechanical Properties of the System Polystyrene-Decalin*†

LESTER D. GRANDINE, JR.,‡ AND JOHN D. FERRY
Department of Chemistry, University of Wisconsin, Madison, Wisconsin
(Received February 16, 1953)

The dynamic rigidities and viscosities of solutions of polystyrene (number-average molecular weight 197 000) in decalin have been measured by three experimental methods over ranges of concentration from 10 to 62 percent, of temperature from −5° to 50°C, and of frequency from 30 to 2000 cps. Similar measurements were made on the undiluted polystyrene from 115° to 133°C and 30 to 2800 cps. The data for the solutions, when reduced to a reference state of unit density and viscosity at a standard temperature, provide single composite curves for dynamic rigidity and viscosity. The distribution function of relaxation times calculated from these curves overlaps that previously derived from stress relaxation measurements on this system, and the combined function covers a range of time scale of ten powers of ten. It shows the three characteristic regions previously identified in the relaxation distribution function of polyisobutylene; there are no obvious anomalies associated with the fact that decalin is a very poor solvent for polystyrene. The data for the undiluted polymer, when reduced to a standard temperature, provide single composite curves for dynamic rigidity and viscosity. The temperature reduction factors provide apparent energies of activation for relaxation which agree with those for viscous flow obtained by Fox and Flory. The relaxation distribution function of the undiluted polymer is similar in shape to, but somewhat sharper than, those for other polymers in the transition from soft to glassy consistency. When reduced to a reference state of unit density and viscosity, the distribution function of the undiluted polymer lies near that reduced from data on solutions, but is considerably sharper than the latter.

INTRODUCTION

DYNAMIC mechanical properties of concentrated solutions of polystyrene,[1] polyisobutylene,[2,3] and polyvinyl acetate[4,5] in several different solvents have been measured in previous investigations. By combining results at different concentrations and temperatures, using the method of reduced variables,[6] the distribution function of relaxation times has been determined in several cases over a time range of several powers of ten. In these studies the solvents have been good to moderate in the sense that the thermodynamic coefficient μ has ranged from 0.38 to perhaps 0.48 (with the exception of some fragmentary data on polyvinyl acetate in diisopropyl ketone,[4] a system which does not mix in all proportions).

The present study was undertaken on solutions of polystyrene in decalin, a poor solvent, with which it does not mix in all proportions below 10°C, to determine whether the mechanical properties would be significantly different in a system on the verge of phase separation. By combining data from three different dynamic methods, together with previously published data on stress relaxation in concentrated solutions following cessation of steady-state flow,[7] the distribution function of relaxation times has been determined over ten decades of logarithmic time—the most extensive range thus far reported for a polymer in solution. In addition, dynamic measurements on pure polystyrene in its softening range have provided the relaxation distribution in this region, for comparison with the behavior of polystyrene in solution as well as with that of other polymers in the solid state.

MATERIALS AND METHODS

The polystyrene was a specially prepared sample, numbered 19F, which was kindly given us by Mr. R. F. Boyer, Dow Chemical Company. Its weight average molecular weight was 370 000, determined by light scattering in the manufacturer's laboratories;[8] the number average value was 197 000, determined by osmotic pressure measurements in methyl ethyl ketone in our laboratory. (Linear extrapolation of osmotic pressure measurements at 25°C gave a value of 0.483 for the thermodynamic constant μ; the Flory-Krigbaum extrapolation[9] gave a value of 0.283 dl/g for the virial coefficient Γ_2. The same molecular weight was obtained in both extrapolations).

The decalin was distilled through a 30-plate Oldershaw column and collected over a boiling range of 187° to 194°C. Its density (ρ_1) was measured at several temperatures from 12° to 40° and was found to follow the empirical equation $\rho_1 = 0.8898 - 0.00074(t-12)$, where t is the centigrade temperature. According to the data of Seyer and Walker,[10] this corresponds to a mixture of 47 percent *trans*- and 53 percent *cis*-isomers.

The poor solvent power of decalin for polystyrene is evidenced by formation of two phases at lower tem-

* Part XI of a series on Mechanical Properties of Substances of High Molecular Weight.

† Presented at the Annual Meeting of the Society of Rheology, Philadelphia, 1952.

‡ Union Carbide and Carbon Fellow in Physical Chemistry, 1950–1952. Present address: E. I. du Pont de Nemours and Company, Wilmington, Delaware.

[1] J. D. Ferry, J. Am. Chem. Soc. **64**, 1323 (1942).

[2] J. N. Ashworth and J. D. Ferry, J. Am. Chem. Soc. **71**, 622 (1949).

[3] Ferry, Fitzgerald, Johnson, and Grandine, J. Appl. Phys. **22**, 717 (1951).

[4] W. M. Sawyer and J. D. Ferry, J. Am. Chem. Soc. **72**, 5030 (1950).

[5] Ferry, Sawyer, Browning, and Groth, J. Appl. Phys., **21**, 513 (1950).

[6] J. D. Ferry, J. Am. Chem. Soc. **72**, 3746 (1950).

[7] Schremp, Ferry, and Evans, J. Appl. Phys. **22**, 711 (1951).

[8] D. J. Streeter and R. F. Boyer, Ind. Eng. Chem. **43**, 1790 (1951).

[9] P. J. Flory, and W. R. Krigbaum, J. Chem. Phys. **18**, 1086 (1950).

[10] W. F. Seyer and R. D. Walker, J. Am. Chem. Soc. **60**, 2125 (1938).

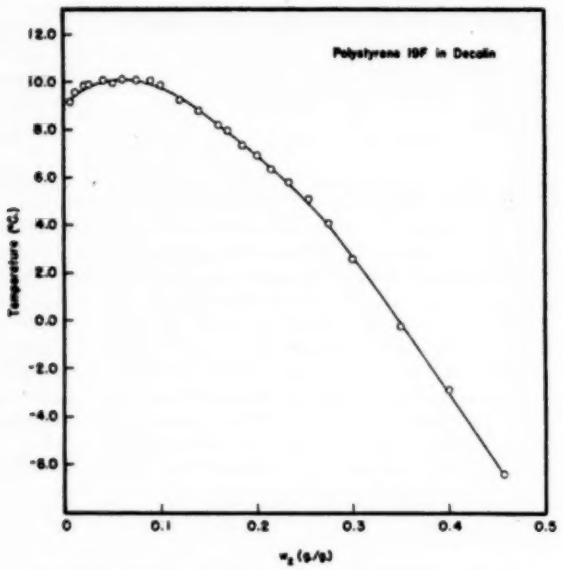


FIG. 1. Temperature of formation of two phases plotted against weight fraction of polystyrene in decalin.

peratures. At high concentration the phases do not actually separate, but the solution becomes very opaque and there is a large increase in rigidity. The temperature at which this phenomenon appears is plotted against concentration in Fig. 1. The critical miscibility temperature is 10.2°C. If these data are used to calculate the values of μ at various temperatures by the lattice theory,[11] the results conform to the empirical equation $\mu = -2.42 + 833/T$, where T is the absolute temperature. (This gives a value of 0.38 at 25°, considerably higher than that obtained in very dilute solution.) The mechanical measurements on solutions were made over a range from −5° to 50°C, corresponding to μ-values of 0.69 to 0.16, according to the above equation. No measurements were made at or below the precipitation temperature at any concentration; but the 25 percent solution was studied at temperatures down to within 5° of this point.

The poor solvent power of decalin is also shown by the abnormally low intrinsic viscosity of polystyrene 19F in this solvent, which we determined as 0.570 dl/g, in agreement with the figure of 0.564 given by Streeter and Boyer.[8] Their values in different solvents range from 0.50 for the poorest to 1.36 for the best.

It is of interest that the curve of Fig. 1 is very similar to the corresponding plot of Fox and Flory[12] for polystyrene in cyclohexane, a chemically similar solvent.

Each concentrated solution was made up by weight, allowing polymer and solvent to remain in contact until the system was homogeneous, without milling or other violent disturbance. To test the possibility of degradation during the rather long periods of storage involved, a 30 percent solution was diluted after 5 months, and an intrinsic viscosity was determined; the value was identical with that obtained from a freshly prepared dilute solution.

The dynamic viscosities, η', at concentrations from 10 percent to 20 percent at 25°C and at concentrations from 13 percent to 25 percent at 35°C, were measured by the single transducer Model 2 of Smith, Ferry, and Schremp,[13] covering a frequency range of 200 to 400 cps. It was impossible to get reliable values of the dynamic rigidity G' in this range.

The wave rigidity $\bar{G}$ and damping index λ/x_0 were measured at concentrations from 25 percent to 46 percent, and temperatures from 10° to 50°C, by the method of transverse wave propagation,[2,4] covering a frequency range of 250 to 2000 cps. The densities of the solutions were estimated for these and other calculations by assuming that the volumes of polymer and solvent are additive at 25° and that the thermal expansion coefficients of the solutions are the same as that of the solvent. Values of G' and η' were calculated from $\bar{G}$ and λ/x_0 by the usual formulas.[2,5]

Measurements of G' and η' were made on a 62 percent solution, of gumdrop-like consistency, with the double transducer constructed in this laboratory by Dr. E. R. Fitzgerald.[14] We are much indebted to Dr. Fitzgerald for supervision and assistance in the operation of this apparatus. The temperature and frequency ranges were −5° to 25° and 30 to 1500 cps, respectively. Two pairs of samples were employed, of dimensions $\frac{5}{16}$ in. (diameter) by $\frac{1}{8}$ in. (thickness) and $\frac{11}{16}$ in. by $\frac{1}{16}$ in., respectively. Calculated results for both G' and η', using nominal sample coefficients,[14] were 31 percent lower for the thicker sample; such a discrepancy has been shown to be attributable to bulging,[14] and the results were therefore adjusted by a constant factor, assuming the values for the thinner sample to be

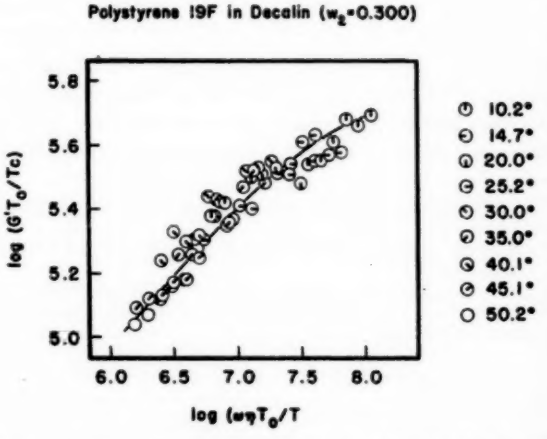


FIG. 2. Reduced dynamic rigidity, in dyne/cm², of a 30.0 percent solution plotted logarithmically against reduced frequency at nine different temperatures identified in margin.

[11] P. J. Flory, J. Chem. Phys. **10**, 51 (1942).
[12] T. G. Fox, Jr., and P. J. Flory, J. Am. Chem. Soc. **73**, 1915 (1951).
[13] Smith, Ferry, and Schremp, J. Appl. Phys. **20**, 144 (1949).
[14] E. R. Fitzgerald and J. D. Ferry, J. Colloid Sci. **8**, 1 (1953).

correct. The sample coefficients at temperatures other than 25° were calculated by taking the volume expansion coefficient of the 62 percent solution as 0.875×10^{-3} deg^{-1}.

Measurements on the pure polymer were also made in the double transducer apparatus, from 115° to 133°C. One pair of samples was employed, $\frac{5}{16}$ in. (diameter) by $\frac{3}{8}$ in. (thickness). The results are believed to be unaffected by bulging, since the samples were inserted at room temperature where the polymer is hard. The volume expansion of the polymer was calculated from data of Fox and Flory.[15] The frequency range in this case was from 30 to 2800 cps.

RESULTS

For brevity,‖ the results on solutions are presented in the form of the reduced variables $G_r' = G'T_0/Tc$ and $\eta_r' = \eta'/\eta$, plotted against the reduced frequency, $\omega_r = \omega\eta T_0/Tc$, representing reduction to a hypothetical reference state of unit density and unit viscosity at a standard temperature.[6] Here T is the absolute temperature of measurement and $T_0 = 298$°K, c is concentration in g/cc, and η is the steady-flow viscosity, obtained by interpolation from data presented elsewhere.[16] According to certain assumptions,[6] the most important of which are that all relaxation times depend identically on temperature and concentration, such plots should yield single curves for G_r' and η_r' at all temperatures and concentrations.

The reduced plots (logarithmic) are illustrated first by G_r' and η_r' for a 30.0 percent solution from wave propagation data at nine different temperatures, in Figs. 2 and 3. In each case a single curve is obtained within the scatter which is characteristic of wave propagation measurements, showing that all relaxation times involved here have the same temperature dependence. Similar curves were obtained for wave propagation data at concentrations of 25.0, 35.0, 40.0, and 45.7 percent. Because of the enormous number of points, it is not practical to show them all on a single graph; instead, a single composite curve was drawn for G_r' and for η_r' at each concentration, and these curves are reproduced together as dashed lines in Fig. 4. The individual points for η_r' from single transducer measurements are also included in Fig. 4. The composite wave propagation curves lie quite near each other. The transducer points lie on a single curve which fits in very well with the wave propagation curves. Heavy curves are drawn through all the data to represent the reduced variables over the concentration range from 10.1 to 45.7 percent.

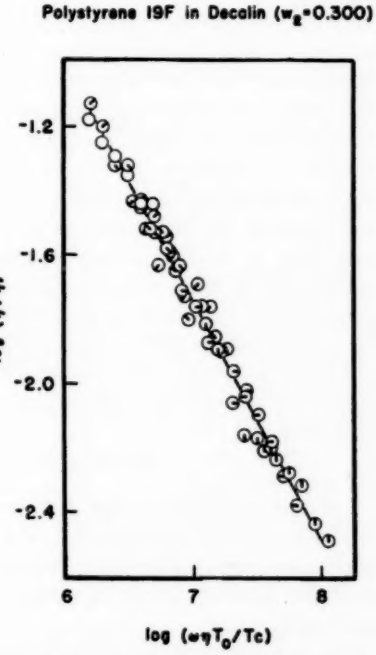


FIG. 3. Reduced dynamic viscosity of a 30.0 percent solution plotted logarithmically against reduced frequency at nine different temperatures (key same as in Fig. 2).

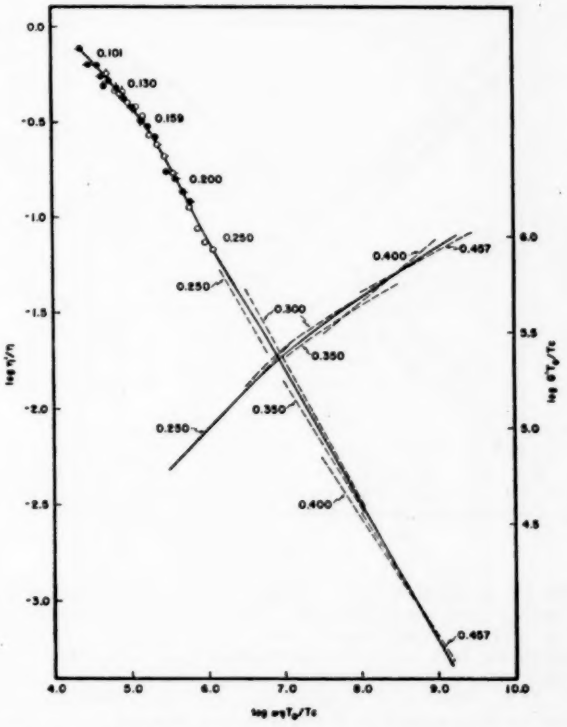


FIG. 4. Reduced dynamic rigidity and viscosity of solutions from 10.1 to 45.7 percent polymer. Wave propagation data plotted as composite curves (dashed) for different concentrations; single transducer data plotted as individual points (solid circles, 25°; open circles, 35°C). Reference state unit density and unit viscosity. Numbers denote concentration as weight fraction.

The data from the double transducer on the 62 percent solution are presented separately in Fig. 5. The sequence of temperatures in these measurements was 25°, 0°, −5°, 10°, 5°, 0°, 15°, 20°, 0°, 25°. The three sets of data at 0° showed a progressive increase in G' of about 7 percent for each repetition, and the final values at 25° were considerably higher than the

[15] T. G. Fox, Jr., and P. J. Flory, J. Appl. Phys. 21, 581 (1950).
‖ The complete data are given in the Ph.D. thesis of L. D. Grandine, Jr., University of Wisconsin, 1952.
[16] Ferry, Grandine, and Udy, J. Colloid Sci. (to be published).

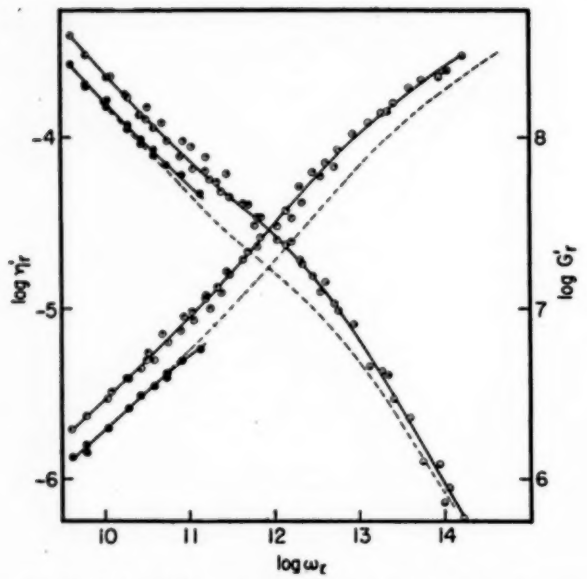


FIG. 5. Reduced dynamic rigidity and viscosity of 62 percent solution plotted logarithmically against reduced frequency at seven different temperatures. Solid circles, first run at 25°C; other circles, later runs at temperatures as follows: downward pip, −4.6°; successive 45° rotations counterclockwise, 0.3°, 4.9°, 9.9°, 14.8°, 19.8°, 25.0°. Dashed line: data from later runs adjusted on the assumption of a steady flow viscosity increase of a factor of 2.5. Reference state unit density and unit viscosity.

initial. These discrepancies are undoubtedly due to loss of a small amount of solvent during the four days required for measurements. However, when plotted in reduced form, all the data superpose to form composite curves except the first set at 25° (Fig. 5). It is believed,

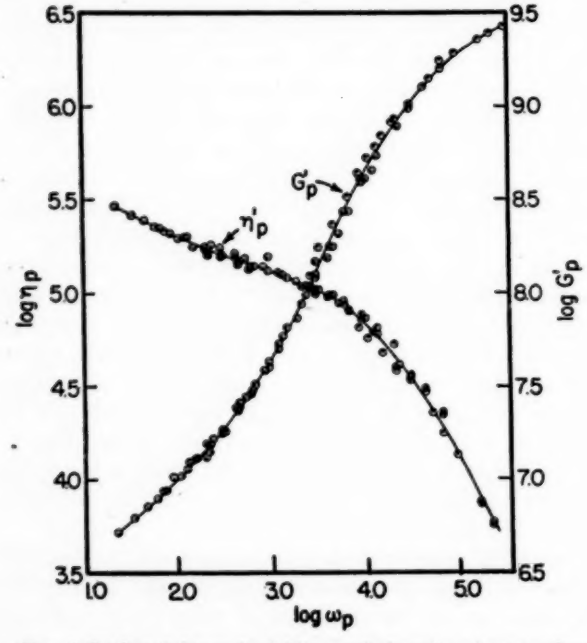


FIG. 6. Reduced dynamic rigidity and viscosity of pure polymer plotted logarithmically against reduced frequency at seven different temperatures (reference state pure polymer at 124.1°C). Upward pip, 114.6°C; successive 45° rotations counterclockwise, 118.2°, 121.2°, 124.1°, 127.0°, 129.8°, 133.2°.

TABLE I. Temperature reduction factors for pure polystyrene, referred to 124.1°C.

Temperature °C	Log a_T
114.6	1.20
118.2	0.69
121.2	0.35
124.1	0
127.0	−0.35
129.8	−0.63
133.2	−0.94

therefore, that the major solvent loss occurred after the first 25° run; the latter corresponds to the original concentration of 62 percent, and the remaining measurements to a slightly higher concentration. The principal effect of a small concentration increase on the reduced variables is a change in steady flow viscosity. By assuming an increase in η by a factor of 2.5 (which would correspond to a concentration increase of about 2 percent), the reduced curves for G_r' and η_r' are shifted to give the dashed curves in Fig. 5, which agree closely with the data from the first 25° run and also fit in very well with the composite curves of Fig. 4.

The data from the double transducer on the pure polymer are presented in the form of the reduced variables G_p' and η_p', plotted against the reduced frequency $\omega_p=\omega a_T$. The subscript p denotes[3] that the reference state is the pure polymer at a standard temperature T_0, which is taken here as 397.2°K, the approximate midpoint of the range covered. Because the range of consistencies covered extends into the glassy region, the reduction treatment is more complicated than previously employed for a pure polymer.[3] The mechanical data are first calculated as the real and imaginary parts of the complex compliance, $J'-iJ''=1/(G'+i\omega\eta')$. These are reduced by the following formulas:[17] $J_p'=J'[T\rho/T_0\rho_0+(J_\infty/J')(1-T\rho/T_0\rho_0)]$ and $J''_p=J''(T\rho/T_0\rho_0)$. Finally $G_p'+i\omega\eta_p'=1/(J_p'-iJ_p'')$. Here J_∞ is the glassy compliance, estimated as 2.2×10^{-10} cm²/dyne; and ρ and ρ_0 are the densities at T and T_0, calculated from data of Fox and Flory.[15] The factor a_T is chosen empirically. With a suitable choice of a_T, both G_p' and η_p' superpose to form composite curves, as shown in Fig. 6. The values of a_T are given in Table I and plotted in Fig. 7, together with values of ΔH_a, the apparent activation energy for relaxation, calculated as $Rd\ \ln a_T/d(1/T)$.

DISCUSSION

The mechanical behavior of the pure polymer, which for simplicity is discussed first, is characterized by two functions—one of time and one of temperature, these variables having been separated by the reduction treatment.

The temperature function is represented by either a_T or ΔH_a, shown in Fig. 7. According to the postulates

[17] J. D. Ferry and E. R. Fitzgerald, J. Colloid Sci. 8, 224 (1953).

TABLE II. Distribution function of relaxation times for pure polystyrene reduced to 124.1°C.

log τ	log Φ (from G') (Φ in dyne/cm^3)	log Φ (from η') (Φ in dyne/cm^3)
−5.0	8.96	9.09
−4.5	8.82	8.92
−4.0	8.60	8.54
−3.5	8.10	8.09
−3.0	7.58	7.40
−2.5	7.16	7.01
−2.0	6.73	6.66
−1.5	6.42	6.32

of the reduction method, it is closely related to the temperature dependence of the steady flow viscosity,[3] and ΔH_a should not differ by more than 0.5 kcal from the apparent activation energy for flow. The latter has been determined by Fox and Flory,[15] who give an empirical formula for it as a function of temperature and molecular weight. Values calculated from their formula are included as a dashed line in Fig. 7 and show gratifying agreement with the apparent activation energy for relaxation. This confirms further that all relaxation times have the same temperature dependence. It also suggests, following the interpretation of Fox and Flory, that the extremely high temperature dependence of relaxation observed in this region is associated, like that of flow, with the rapid relative decrease of free volume as the glass transition is approached. The latter point is about 100°C.[15]

The time function may be represented by Fig. 6 directly, showing the increase of G' from a rubbery to a glassy magnitude with increasing frequency. Alternatively, the distribution function $\Phi d\ln\tau$, the differential contribution to rigidity associated with relaxation times whose logarithms lie in a range $d\ln\tau$, may be employed. This can be calculated from either G' or η' and thus provides an internal test of the accuracy of the experimental data. Values of Φ calculated from Fig. 6 by second approximation formulas[18] are listed in Table II. The agreement between the calculations from G' and η' is satisfactory.

For comparison of Φ for polystyrene with those of other polymers, use is made of the characteristic time K of Tobolsky,[19] defined as the time in a stress relaxation experiment where the log-log plot of stress against time passes through an inflection. From the relations between stress relaxation and G' it follows[20] that this will differ only slightly from the reciprocal of the (angular) frequency at which the log-log plot of G' against frequency passes through an inflection. Graphical differentiation of Fig. 6 shows that the latter is $10^{3.6}$ sec^{-1}, where $d\log G'/d\log\omega$ attains a maximum value of 1.14. Thus $K=10^{-3.6}$ sec at 124.1°C.

Shapes of distribution functions of various polymers

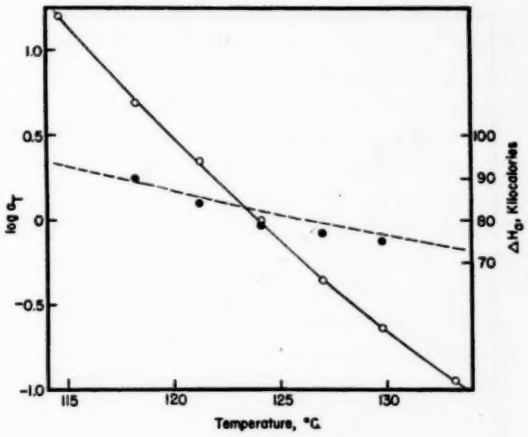


FIG. 7. Reduction factor a_T for pure polymer (open circles) and apparent activation energy for relaxation (solid circles), plotted against temperature. Dashed line: apparent activation energy for viscous flow from the data of Fox and Flory (see reference 15).

may be compared by plotting log Φ against log τ/K, which places them all in the same range of time scale. The values in Table II are presented in this manner in Fig. 8, together with plots for a GRS vulcanizate derived[17] from stress relaxation data of Tobolsky[19] and for polyisobutylene derived from dynamic measurements in this Laboratory.[21] The shapes are quite similar, though the polystyrene curve is somewhat sharper than the others. Additional evidence is thus provided that, in the transition range from soft to glassy consistency, the shape of the distribution does not depend greatly on chemical composition. However, the temperature at which this transition falls in some specified time scale does of course differ widely.

Turning to the behavior of solutions, the variables have been separated in this case into a function of

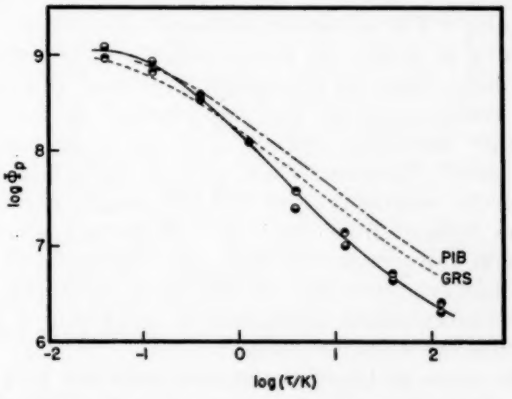


FIG. 8. Distribution function of relaxation times for pure polymer, plotted logarithmically against log τ/K, where K is the characteristic time of Tobolsky. Top shaded, from G'; bottom shaded, from η'. Dashed lines are corresponding functions for a GRS vulcanizate (see reference 19) and a sample of polyisobutylene distributed by the National Bureau of Standards (see reference 21).

[18] J. D. Ferry and M. L. Williams, J. Colloid Sci. 7, 347 (1952).
[19] Bischoff, Catsiff, and Tobolsky, J. Am. Chem. Soc. 74, 3378 (1952).
[20] R. S. Marvin, Phys. Rev. 86, 644 (1952).
[21] Fitzgerald, Grandine, and Ferry, J. Appl. Phys. 24, 650 (1953).

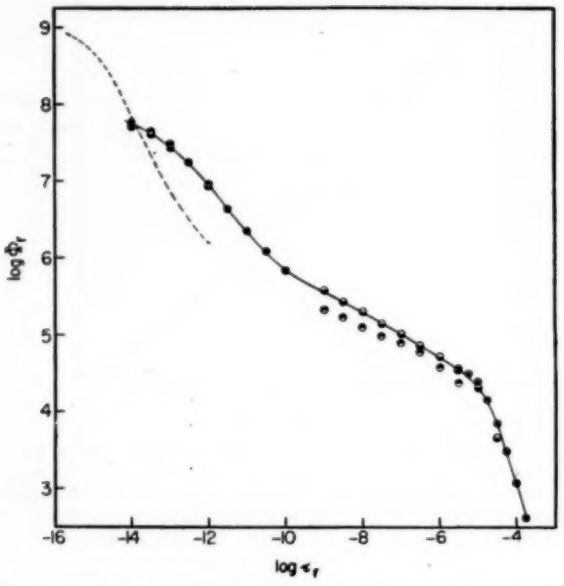


FIG. 9. Distribution function of relaxation times for solutions, plotted logarithmically, reduced to unit density and unit viscosity. Top shaded, from G'; bottom shaded, from η'; crossed circles, data from stress relaxation (see reference 7). Dashed line: distribution function of pure polymer reduced to same reference state.

time and a function of temperature and concentration. The latter has been absorbed into the steady flow viscosity by the reduction treatment, and its analysis would be equivalent to analysis of the dependence of viscosity on concentration and temperature, which is beyond the scope of this paper.

The time function may again be represented by Φ, calculated from both G' and η'. Figure 9 shows values derived from the data of Figs. 4 and 5, as well as values obtained from previous studies of stress relaxation in concentrated solutions by Schremp, Ferry, and Evans,[7] using in all cases the recent second approximation formulas.[18] The agreement between calculations from G' and η' is satisfactory except from the wave propagation data, where the experimental precision is limited. The results from all four experimental methods fit together extremely well. It is of particular interest that stress relaxation values, reduced from data on a 46 percent solution,[7] agree well with single transducer values reduced from data on a 10 percent solution. Thus in this system, at least, the hypothesis of the reduction treatment that all relaxation times have the same concentration dependence is valid over a considerable concentration range.

The curve of Fig. 9 is the most extended that has been obtained thus far for solutions. It shows the three regions which have previously been identified in undiluted polyisobutylene[3]—a steep portion at short times which is the beginning of the transition to glassy consistency, a flatter central portion, and a sharp drop at long times which is associated with the finite lengths of the polymer molecules. The central portion is not nearly so flat nor so extended as in solid polyisobutylene of viscosity average molecular weight 1.1 million; this difference may be attributed partly to the presence of solvent and partly to the considerably lower molecular weight of the polystyrene.

The data for the pure polymer, which in Table II are reduced to a standard temperature of 124.1°C, may be further reduced to the hypothetical reference state of Fig. 9 by the formulas[3] $\Phi_r=\Phi_p(298/397)/\rho$ and $\tau_r=\tau_p\rho/\eta$, where ρ and η are the density and viscosity of the polymer at 124.1°C. The latter values were calculated from data of Fox and Flory[15] to be 1.020 g/cc and $10^{10.70}$ poises. The Φ-values of Table II thus reduced to unit density and viscosity at 25°C are included in Fig. 9 as a dashed line. They do not coincide with the values reduced from solution data (which in this region are provided primarily by the 62 percent solution); the distribution function is somewhat sharper in the undiluted polymer than in solution. Thus, the concentration dependence of all relaxation times is not the same between 62 and 100 percent.

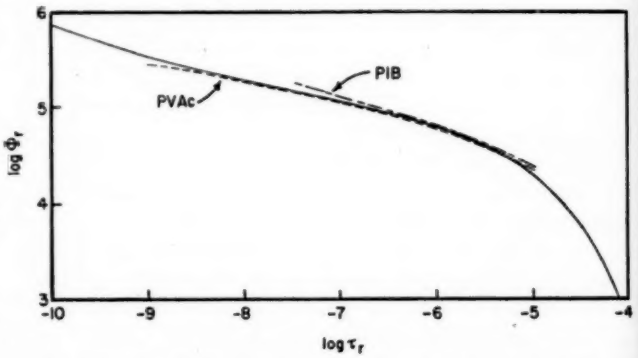


FIG. 10. Distribution function of relaxation times from Fig. 9 (solid line) compared with similar functions for polyvinyl acetate (PVAc) in 1, 2, 3-trichloropropane and polyisobutylene (PIB) in xylene, all polymers having roughly similar molecular weights.

Nevertheless, the two curves fall in the same region of time scale. As concluded previously from comparison of Φ-functions for polyisobutylene and its solutions on this basis,[3] the reduction treatment is of value even here in bringing the data close enough together to facilitate interpolation, as well as aiding interpretation.

Finally, it is of interest to compare the shape of the distribution function in this rather poorly compatible polymer-solvent system with those previously determined for other polymers of similar molecular weights in good solvents. There are no other data covering such a wide range of time scale, but the central portion of Fig. 9 can be compared with similar functions for unfractionated polyvinyl acetate[5,18] of number-average molecular weight 140 000 in 1,2,3-trichloropropane ($\mu=0.38$), and for a roughly fractionated polyisobutylene[22] of viscosity-average molecular weight 320 000 in xylene. These functions, each reduced to a reference

[22] M. F. Johnson, Ph.D. thesis, University of Wisconsin, 1950.

state of unit density and unit viscosity, are plotted in Fig. 10. They are practically identical. This close agreement is no doubt fortuitous, since some influence of chemical constitution is to be expected; nevertheless, it is evident that the mismatch of cohesive energy densities of the system polystyrene-decalin has no obvious effect on dynamic mechanical properties in this range of time scale, as long as the system exists as a single phase.

ACKNOWLEDGMENTS

This work was supported in part by a grant from Research Corporation, and in part by the Research Committee of the Graduate School of the University of Wisconsin from funds supplied by the Wisconsin Alumni Research Foundation. We are indebted to Union Carbide and Carbon Corporation for support through its Fellowship in Physical Chemistry. Since March, 1951 these investigations have been part of a program of research on the physical structure and properties of cellulose derivatives and other polymers supported by the Allegany Ballistics Laboratory, Cumberland, Maryland, an establishment owned by the United States Navy and operated by the Hercules Powder Company under Contract NOrd 10431. The assistance of Miss Martha Maybury in calculations is gratefully acknowledged.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Mechanical Investigations of Elastomers in a Wide Range of Frequencies*

WLADIMIR PHILIPPOFF
The Franklin Institute Laboratories for Research and Development, Philadelphia, Pennsylvania
(Received March 23, 1953)

A method has been found for measuring the shear moduli and losses of elastomers in a range of frequencies of about 1 to 10^6, from 10 cps down. This method provides for large cyclic shear stresses and deformations without employing resonance methods. Only one sample is needed for a complete study, the frequency of stress being the only variable. The method gives results that fit the "box-distribution function" developed by Kuhn and Kuenzle. Results for a series of elastomers show that such investigations are important in ascertaining the mechanical properties of elastomers in practice.

I. INTRODUCTION

MANY experimental investigations have shown that plastics in general, as well as elastomers and their solutions, do not behave like Newtonian liquids. They both have viscosity and elasticity, the combination of which, according to the Maxwell postulate, causes their properties to be dependent on time, and leads us to call them "viscoelastic bodies." This dependence does not exist either for the Newtonian liquid whose mechanical properties are determined by its viscosity, for the elastic body whose properties are determined by its elasticity. The time element introduces a peculiar difficulty in measuring the mechanical properties of these substances. Whereas the well-known Maxwell body describes this time dependence in simple mathematical expressions, it does not apply to the experimental results. The properties of viscoelastic bodies are not described by a Maxwellian body, but rather by a distribution of an infinite number of Maxwellian bodies, each having a so-called "relaxation time."

The existence of this infinite distribution calls imperatively for an investigation of the properties of these viscoelastic bodies in as wide a range of frequencies or times as possible in order to define them. Most investigators have used one method in a range of frequencies no wider than about 1 to 100. This range can be extended by using the observation of Ferry[1] that the properties of viscoelastic bodies investigated at different temperatures, in the case of dynamic investigations with sinusoidal vibrations, seem to be described by shifting one standard curve on the frequency axis. One can superimpose measurements at different temperatures and extend the range of frequencies to about 1 to 1 000 000. But using Ferry's method of "reduced variables" necessitates some assumptions that, though very probable, have as yet not been proven exactly by experiments. It therefore seemed important to have a method for investigating the mechanical properties of viscoelastic bodies in a frequency range of about 1 to 1 000 000 under isothermal conditions. This paper describes a basic method and the results obtained for several elastomers.

II. PRINCIPLES OF OPERATION

In order to measure the viscosity and elasticity of viscoelastic bodies in a wide range of frequencies, it is profitable to use an arrangement that will allow the frequency alone to be changed, while using one sample throughout the investigation. It is well known that the

* Presented before the 24th Annual Meeting of the Society of Rheology, October 30–31, 1952, at The Franklin Institute in Philadelphia, Pennsylvania.

[1] Ferry, Fitzgerald, Grandine, and Williams, Ind. Eng. Chem. 44, 703 (1952).

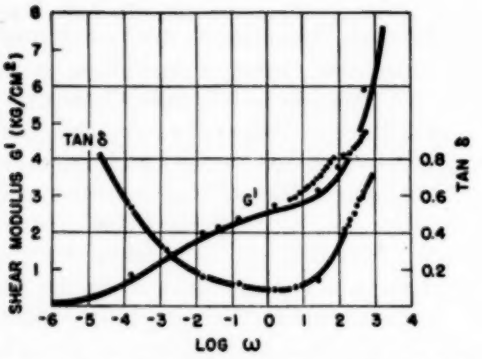


FIG. 1. Frequency dependence of G' and tanδ for polyisobutylene at 25°C. G' is the dynamic shear modulus. δ is the loss angle.

size of the sample affects the pure geometry of mechanical strains for higher degrees of deformation and that the geometry is not always exactly known. Using a single specimen, we at least know that the geometric corrections are constant throughout the whole investigation, even if they influence the absolute values measured. This degree of incertitude is not extremely important; the mechanical constants of viscoelastic bodies always vary within a certain range due already to some variation in composition or molecular weight and cannot be considered as absolute physical constants of the substance like density, etc.

At the start of this investigation it was deemed important to work with forces and deformations not infinitely small, but in the range of those encountered in practice. This is also important should one want to investigate the nonlinear mechanical properties of viscoelastic bodies. Simple shear gives the most linear relationships of all the methods of stressing, especially with elastomers which allow a high degree of elongation. Furthermore, stresses in simple shear are nearly uniform throughout the sample.

To investigate elastomers in a wide range of frequencies, we can use the stress-relaxation method or an investigation under cyclical vibrations. The stress-relaxation method as used by Tobolsky[2] and his associates allows a practically infinite range of times, but it gives only one quantity: the time-dependent modulus of elasticity. An investigation with cyclical vibrations allows us to measure two quantities: the dynamic modulus of elasticity, and a quantity proportional to the losses of the system. This latter can be described as the viscosity, η, or the imaginary part of the shear modulus of elasticity, $\eta \times \omega$, where ω is the circular frequency of the vibrations. In the case of deformational shear vibrations, the quantities are the dynamic shear modulus G' and the dynamic loss modulus $G'' = \omega\eta$. Using vibrations in a wide range of frequencies practically eliminates resonance methods, because the resonant frequency depends on the square of the mass for a constant sample, and a frequency variation of 1 to 1 000 000 would require a variation of the mass of 1 to 10^{-12}. We therefore selected a nonresonant method. From the discussion of reduced variables, it is clear that in the investigation of elastomers it is not the value of the frequencies, but the range of frequencies that is important.

These basic principles of operation are (a) large range of frequencies, (b) large deformations or forces, and (c) nonresonant method—lead us to the method of the hysteresis loop, i.e., the dependence of deformation on the force for harmonic shear vibrations. The method has been used by several investigators beginning with Schmidt[3] in 1923.

A. Sample

The sample consists of two 1 in.$\times$1 in.$\times\frac{1}{4}$ in. pieces of the material to be investigated, sandwiched between three brass plates $1\frac{1}{4}$ in.$\times 1\frac{1}{2}$ in.$\times\frac{3}{8}$ in. The bond between the sample and the plates, in the case of filled rubber, is formed by vulcanization. In the case of soft

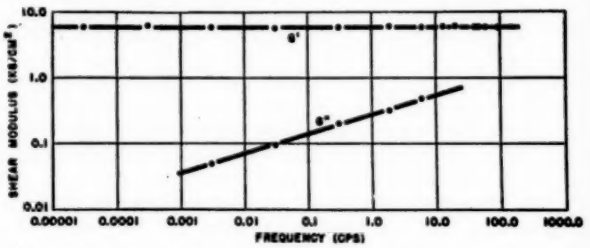


FIG. 2. Frequency dependence of shear modulus, natural gum rubber. G' is the dynamic shear modulus. G'' is the dynamic shear loss modulus.

rubber, polyvinylchloride and Thiokol, the adhesion is so poor that a suitable cement had to be used. The corrections to the calculated values for the finite size of the sample, using formulas for an infinite sample, are about 5 percent according to Read.[4] The maximum double amplitude used was about 1000 microns, or about 8 percent maximum deformation.

The middle plate of the sample sandwich was mounted in a frame connected through a dynamometer to a reciprocating part which gave the sinusoidal motion, whereas the two other plates were solidly clamped to the machine. It is obviously a stressing under simple shear.

B. Method of Stressing

Stresses are provided by an adjustable double cam with a throw between 0 and 4000 microns double amplitude, which could be driven between 13 and 0.000006 cps, giving a range of slightly over 1 to 2 000 000 in frequency. Beyond 13 cps, the system began to be mechanically unstable (unwanted resonant fre-

[2] Tobolsky, Prettyman, and Dillon, J. Appl. Phys. **15**, 380 (1944).

[3] E. Schmidt, Gesundh. In. **46**, 61 (1923).

[4] W. T. Read, J. Appl. Mech. **17**, 349 (1950).

quencies being excited), whereas the lower frequencies required such long times as to be inconvenient. All the bearings were sleeve bearings, rather than ball or roller bearings, to eliminate any possible periodic disturbance. The sample was placed in a liquid bath that could be electrically heated up to 75°C or cooled in a cold room to −30°C. Results reported here were obtained at room temperature (about 25°C or 77°F) without using the temperature control.

C. Sensitive Elements

The force was measured by a dynamometer between the driving member and the frame carrying the sample. The deformation of a steel ring was measured by a Schaevitz (type 040) linear-variable differential transformer. The movement of the frame was picked up by a second linear-variable differential transformer, type 080. The range of forces with one dynamometer was between a few grams and 10 kilograms double amplitude. The range of deformations was between 10 and 1000 microns double amplitude. This gives a very great range of measurable viscosities, amounting to about 1 to 1 000 000, i.e., from 10^6 to 10^{12} poises.

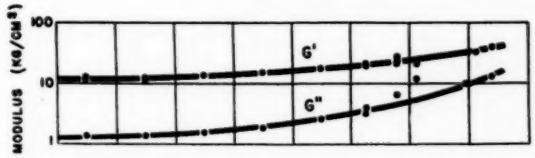


FIG. 3. Natural rubber—tread stock. G' is the dynamic shear modulus. G'' is the dynamic shear loss modulus.

D. Electronic Equipment

Both differential transformers were excited by a 4 volt, 1600 cycle, stabilized frequency generator. The output of each differential transformer after passing the gain control was amplified and rectified through a phase sentitive bridge and low-frequency bypass filter in the usual manner. Both circuits were entirely symmetrical so that the electronics did not introduce any amplitude or phase errors in the range of frequencies investigated. The outputs of the filters were fed to the x and y plates of a Du Mont 304-*H* oscilloscope and photographed by a Du Mont Land camera. The amplitudes of the axes were regulated by the gain control, using the amplifiers at constant output, so that the trace could be well measurable on the oscilloscope screen. Absolute measurements were made possible by using a switch step potentiometer (attenuator) excited by the same voltage as the transformers, which was directly calibrated in kilograms and microns by putting on loads or motions in the instrument and comparing them to the electrical amplitudes determined by the attenuator. Motion at various frequencies was produced by a synchronous motor of 3600 rpm, which drove the cam through a series of pulleys and worm-gear boxes. The low frequencies were calculated from the known gear ratios and the higher frequencies were measured by stop watch and revolution counter.

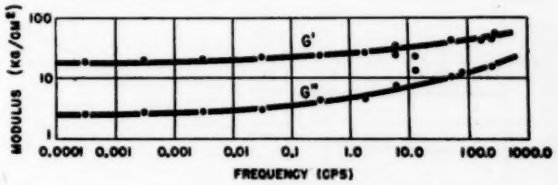


FIG. 4. Cold rubber—tread stock. G' is the dynamic shear modulus. G'' is the dynamic shear loss modulus.

E. General Assembly

The specimen holder with the cam and the sensitive elements, together with the temperature bath, were mounted on a 1600 lb cement block which itself was isolated by steel springs from a concrete pier sunk into the ground. The electronic equipment, the worm drives and other gears were shock-mounted on a table and only connected with the instrument by a rubber V-belt, thus eliminating the transmission of disturbing frequencies during the measurement as far as possible. This was important, especially in measuring over long times or measuring very stiff substances with only a few microns double amplitude or very low forces below 100 grams. The temperature bath was heated by an immersion heating element and cooled by a copper coil with water.

III. RESULTS

In Figs. 1 through 4, the results for a number of elastomers are given. Figure 1 summarizes the results for the National Bureau of Standards' sample of polyisobutylene that had been investigated by a number of laboratories. Our values amplify the frequency range by about two decades at the lower end, and especially important is the strongly defined minimum of the loss angle *vs* frequency. The values are concordant with the ones obtained by stress relaxation or by other methods. The rise of the loss angle with diminishing frequency indicates that at some still lower frequency the polyisobutylene will behave as a Newtonian liquid, i.e., it will flow with a constant viscosity.

In Fig. 2 the measurements on vulcanized rubber gum are shown. The values obtained by A. W. Nolle[5] at high frequencies for a similar composition of rubber gum are concordant, so that with the combination of both methods a continuous range of values from 0.000006 to about 5000 cycles is available. The moduli of elasticity measured by Nolle were divided by three to be compared with the shear modulus measured in this work. The losses of gum rubber are very similar to these of Nolle, except for the absence of a sharp downward trend that he obtained, which was probably at the limit of application of his methods. The losses at low frequencies, below 0.003 cps, could not be measured; the gum rubber

[5] A. W. Nolle, J. Polymer Sci. **5**, 1 (1950).

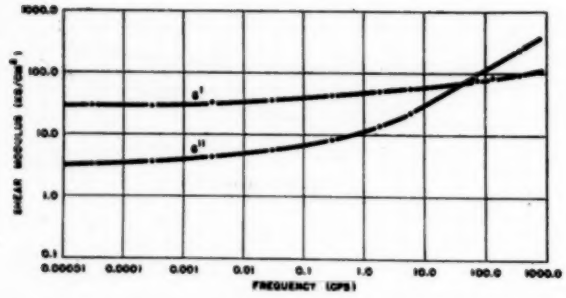


FIG. 5. Frequency dependence of shear modulus, Thiokol *H*-5. G' is the dynamic shear modulus. G'' is the dynamic shear loss modulus.

below these frequencies behaved like an ideal elastic body with losses well below 1 percent.

Figures 3 and 4 show the results for a natural rubber tread and a cold rubber tread stock, made by the Rubber Department of the National Bureau of Standards. Both the modulus and the losses are frequency-dependent. Values in the higher frequency range were obtained by putting our specimen into a Gehman[6] tester where it was stressed in exactly the same way: simple shear. Figure 5 presents the results for a blend of Thiokol obtained from the Thiokol Corporation in Trenton, New Jersey (type *H*-5) and shows at 25°C essentially the same behavior as other elastomers. At higher temperatures Tobolsky and Stern[7] have obtained a behavior very similar to that of Maxwellian bodies. This is definitely not the case at 25°C.

Measurements of the properties of the viscoelastic bodies at different amplitudes of vibration showed that apparently the amplitude does not influence the viscosity to any measurable extent beyond the inherent error limit of the machine. The modulus G' and $\tan\delta = G''/G'$ have possibly a small amplitude dependence, resulting in the modulus being smaller at higher amplitudes and $\tan\delta$ increasing, but the effects are comparatively very small, especially considering the influence of frequency on these properties. It is interesting to note that if viscosity, which in itself is frequency dependent, does not appreciably depend on amplitude, then it is *not the rate of shear* that causes changes but rather the *frequency* that is the most important factor in determining the viscosity of elastomers.

IV. DISCUSSION

All the experimental curves except the one for the shear modulus of natural gum rubber are frequency dependent. They are extremely flat in a frequency range of at least $1:10^6$. This substantiates very well the difficulties of investigating viscoelastic bodies mentioned in the introduction. For similar curves, Kuhn and Kuenzle[8] in 1946 developed a theory that was subsequently called "box distribution"; they assumed that the probability of a partial shear modulus $dG'/d\tau$ for a mechanism with a relaxation time τ was

$$dG'/d\ln\tau = b/a^2 = \text{const.}$$

This idea, after adequate mathematical treatment modifying somewhat the original formulas, led to the following expressions for the frequency dependence of dynamic modulus G' and G'' and the loss angle $\cot\delta = G'/G''$

$$\frac{1}{G'} = \frac{a-\ln\omega}{b}, \tag{1}$$

$$1/G'' = (2/\pi b)(a-\ln\omega)^2, \tag{2}$$

$$\cot\delta = (2/\pi)(a-\ln\omega). \tag{3}$$

These formulas are valid under the simplifying assumption

$$X = \frac{\pi^2}{(a-\ln\omega)^2} \ll 1.$$

In Kuhn's case this was true to a large extent. But in some of our samples this was not the case. We have therefore calculated the second approximation of the formulas for values $0 < x < 1$. The first approximation is valid until $x \leqq 0.3$ with an error of 10 percent. For higher values of "x" the $1/G'$ values are *higher* than the ones of the first approximation.

Physically, "a" determines a "high frequency cut-off" of the distribution function, eliminating any shorter times above the frequency f_0 or time τ_0:

$$[a = \ln 2\pi f_0 = \ln(1/\tau_0)]$$

Measurements of Kuhn and Kuenzle in a very limited range of frequencies indicated that such a relationship could exist. Our measurement in a much wider range of frequencies can give a closer check of these relations. By plotting the reciprocal dynamic shear modulus over $\log f$, one should expect, according to (1), linear relationship. As is seen in Fig. 6, this is indeed the case for all the elastomers investigated. One has therefore to consider the assumptions made by Kuhn as to the distribution of the relaxation times to be valid for these elastomers in an extremely wide range of frequencies.

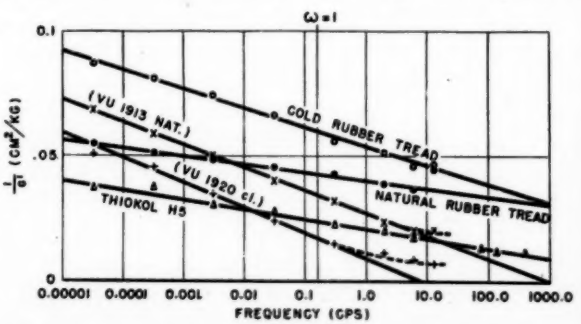


FIG. 6. Plot of $1/G'$ *vs* $\log f$ to test the Kuhn equations.

[6] Gehman, Woodford, and Stambaugh, Ind. Eng. Chem. **33**, 1032 (1941).
[7] A. V. Tobolsky and M. D. Stern, J. Chem. Phys. **14**, 93 (1946).
[8] Kuhn, Kuenzle, and Preissmann, Helv. Chim. Acta **30**, 307, 464 (1947).

TABLE I.

Sample	$\frac{1}{G'}$ at $\omega=1$ (cm²/kg)	$\frac{1}{G'}$ at 10^{-5} cps (cm²/kg)	b (kg/cm²)	a —	τ_0 (sec.)	f_0 (cps)
Natural rubber tread	0.0430	0.0570	691.2	29.71	1.25×10^{-13}	1.628×10^{12}
Cold rubber tread	0.0603	0.0928	267.0	16.02	1.100×10^{-7}	1.151×10^{6}
Thiokol	0.0240	0.0405	586.5	14.06	7.82×10^{-7}	1.943×10^{5}
Polyvinylchloride *VU* 1913 natural	0.0345	0.0733	250.0	8.625	1.8×10^{-4}	8.95×10^{2}
Polyvinylchloride *VU* 1920 clear	0.0173	0.0600	226.0	3.91	2.0×10^{-2}	7.95

This is *not* the case for the polyisobutylene which at very low frequencies approaches the behavior of a liquid.

The validity of the relationships of Kuhn leads to unexpected results. Under no condition can the "true value" of the "static" shear modulus of an elastomer be found, as the dependence on the logarithm of the frequency does not warrant a limiting value at any frequency. There are as yet no indications that this statement could be modified for a range of frequencies above or below the one investigated. Of course some limitation as to the limits of the validity of Eqs. (1) to (3) should exist, but in any case in the lower frequency range no deviations have been found. At the lowest frequencies investigated, the elastomers behave neither as pure elastic bodies nor Newtonian liquids. The viscosity as calculated from G'' has roughly the hyperbolic dependence on frequency as expected by Kuhn. There is no particular value of viscosity and elasticity that could be assigned to the elastomer except perhaps the one at $\omega=1$ ($f=0.159$ cps or $\ln\omega=0$). For Kuhn's relationships to be valid further requires that the properties depend on only two constants, "a," and "b," that occur in different combinations in all the formulas. It is essentially immaterial which particular method one uses in order to determine these constants. This fact is very important as it makes it possible to describe the viscoelastic properties of elastomers, for which the Kuhn relations hold, by only two constants. A tabulation of these constants for our elastomers is given in Table I.

The practical value of the results obtained by this investigation is obvious. Because the mechanical constants of elastomers depend so much upon the frequency of stress, it is clear that one has to measure these constants at the same frequencies that occur in practice; i.e., if the elastomer is to be used in shock absorbers the elastic constants have to be measured at frequencies between, roughly 3 and 50 cps. Should the application be in gaskets under steady load, measurements at very low frequencies such as 1 cycle per day are necessary. The susceptibility of the mechanical properties to changing frequency are different for each elastomer, and range from zero in the case of natural gum rubber up to a change of about 1 to 10 for the polyvinyl product. Thus it is not possible to use a constant factor with which the value of (say) shear modulus at any one frequency should be multiplied in order to get its value at any other frequency. What the actual conditions are can only be ascertained by expanding these investigations to get results for a larger number of different plastics, probably also using different temperatures.

Viscoelastic Properties of Dilute Polymer Solutions

P. E. Rouse, Jr., and K. Sittel
The Franklin Institute, Laboratories for Research and Development, Philadelphia, Pennsylvania
(Received March 23, 1953)

The viscoelastic properties of dilute solutions of chain polymers have been measured at frequencies in the range from 200 cps to 60 kc. The data are compared with theoretical curves calculated from the steady-flow viscosities of the solution and solvent, the molecular weight and concentration of the polymer, and the absolute temperature; this calculation made use of equations obtained from a recently developed theory. The equations contain no adjustable constants. The theory is shown to be at least a good first approximation to the viscoelastic properties of the solutions studied.

The agreement of the data and the theory implies that, up to 60 kc at least, the viscoelastic properties of dilute solutions of polystyrene and polyisobutylene are the result of the thermal motions of the segments of the polymer molecules. These segmental motions coordinate with one another to produce changes in the configurations of the molecules. The configurational changes have associated with them a series of relaxation times. These relaxation times are shown to depend upon molecular weight of the polymer, concentration, type of solvent, and viscosity of solvent.

INTRODUCTION

THE viscoelastic properties of solutions of polymeric molecules have been the subject of many experimental investigations.[1] However, until recently the lack of highly sensitive techniques has limited these experimental studies to concentrations so high that contacts between the polymer molecules strongly affected the thermal motions of the chains. The dependence of these thermal motions, and, consequently, of the viscoelasticity, upon the length and flexibility of the polymer molecules can best be studied by measurements on solutions so dilute that intermolecular contacts are relatively infrequent. Two techniques sufficiently sensitive for such work have recently been developed.[2,3]

Baker and his co-workers[4] have published measurements made by one of these techniques on dilute solutions of several coiling polymers. At the time of their publication a quantitative theory which could be compared with such data was not available. A completely general theory of the viscoelastic properties of solutions of linear polymers had been developed by Kirkwood;[5] unfortunately, the results of his theory are not in a form which can be used for calculation. A treatment by W. Kuhn and H. Kuhn[6] had ignored the fact that under dynamic conditions the configurations of a polymer molecule cannot be adequately described by merely specifying the end-to-end separation. As a result their theory predicted only a single configurational relaxation time. Consequently, Baker[4] expressed his data in terms of a modified Maxwell element (spring and dashpot combination in parallel with a second dashpot). The constants of the dashpots and the spring were calculated from the experimental data obtained at each frequency. It was found that the calculated values were functions of the frequency of measurement. Only by use of a model consisting of several elements could data for more than one frequency be fitted. This result emphasizes the fact that more than one relaxation process is involved in the viscous and elastic behavior of solutions of coiling polymers.

Recently a theoretical calculation[7] has been carried out which produced results that can be compared with experiment. This theory predicts a series of discrete relaxation times associated with the changes of configuration of the polymer molecules. Equations were obtained by means of which the relaxation times and the components of the complex viscosity[8] (or complex rigidity)[9] can be calculated from the steady-flow viscosities of the solution and the solvent, the molecular weight and concentration of the polymer, and the absolute temperature. These equations contain no adjustable constants.

Data are presented here for the viscoelastic properties of dilute solutions of polystyrene in toluene and of polyisobutylene in various solvents. The measurements were made at frequencies in the range from 220 cycles up to 60 kc. The data are compared with curves calculated from the theory.[7] The agreement indicates that the equations obtained from the theory provide a good first approximation to the viscoelastic properties of the solutions studied.

SUMMARY OF THEORY

The theory[7] referred to above relates the elastic and viscous properties of solutions of polymers to the dis-

[1] W. Philippoff, Physik. Z. **35**, 884, 900 (1934); J. M. Kendall, Rheol. Bull. **12**, 26 (1941); J. D. Ferry, J. Am. Chem. Soc. **64**, 1323, 1330 (1942); J. R. Van Wazer and H. Goldberg, J. Appl. Phys. **18**, 207 (1947); Ferry, Sawyer, and Ashworth, J. Polymer Sci. **2**, 593 (1947); J. N. Ashworth and J. D. Ferry, J. Am. Chem. Soc. **71**, 622 (1949); Smith, Ferry, and Schremp, J. Appl. Phys. **20**, 144 (1949); Ferry, Sawyer, Browning, and Groth, J. Appl. Phys. **21**, 513 (1950); W. M. Sawyer and J. D. Ferry, J. Am. Chem. Soc. **72**, 5030 (1950); Schremp, Ferry, and Evans, J. Appl. Phys. **22**, 711 (1951).
[2] W. P. Mason, Trans. Am. Soc. Mech. Eng. **69**, 359 (1947).
[3] Sittel, Rouse, and Bailey (to be published).
[4] Baker, Mason, and Heiss, J. Polymer Sci. **8**, 129 (1952).
[5] J. G. Kirkwood, Rec. trav. chim. **68**, 649 (1949).
[6] W. Kuhn and H. Kuhn, J. Colloid Sci. **3**, 11 (1948).
[7] P. E. Rouse, Jr., J. Chem. Phys. (to be published).
[8] A. Gemant, Trans. Faraday Soc. **31**, 1582 (1935).
[9] R. L. Wegel and H. Walther, Physics **6**, 141 (1935).

turbance of the distribution of configurations of the flexible polymer molecules. This disturbance increases the free energy of the system. The thermal motions of the segments of the molecules cause the distribution of configurations of the molecules to drift continuously toward its equilibrium form, dissipating the excess free energy. In this calculation, a polymer molecule is considered as being made up of N equal submolecules, each of which is a portion of polymer chain just long enough so that the separation of its ends obeys, to a first approximation, a Gaussian probability function. The necessary coordination of the motions of the various parts of a molecule is taken into account by introducing the requirement that the motion of the end of a submolecule changes the configurations both of that submolecule and of the adjoining submolecule.

The response of a solution of a polymer to a sinusoidally varying shearing stress can be expressed in terms of a complex viscosity[8] $\eta^*=\eta_1-i\eta_2$ or a complex rigidity[9] $G^*=G_1+iG_2$. The expressions for these quantities produced by the calculation[7] are given in the following equations:

$$\eta_1=\frac{G_2}{\omega}=\eta_s+nkT\sum_{p=1}^{N}\frac{\tau_p}{1+\omega^2\tau_p^2}, \tag{1}$$

$$\eta_2=\frac{G_1}{\omega}=\frac{nkT}{\omega}\sum_{p=1}^{N}\frac{\omega^2\tau_p^2}{1+\omega^2\tau_p^2}, \tag{2}$$

$$\tau_p=\frac{6(\eta_o-\eta_s)}{\pi^2p^2nkT}. \tag{3}$$

In these equations ω=angular frequency of applied shearing stress. η_s=viscosity of solvent, n=number of polymer molecules per unit volume of solution, k=Boltzmann's constant, T=absolute temperature, and η_o=viscosity of the solution in steady flow.

The result of the calculation indicates that the processes involved are characterized by a series of relaxation times τ_p. The pth process makes a contribution of $nkT\tau_p$ to the steady-flow viscosity and a contribution of nkT to the instantaneous rigidity of the solution.

In the range of frequencies for which $\omega\tau_1<N^2/250$, the values calculated from these equations are insensitive to increases in N. The summations in Eqs. (1) and (2) have therefore been evaluated under the assumption that N is infinite. The theoretical values given in the tables and the curves in the figures were calculated by means of Eqs. (1), (2), and (3) with N set equal to infinity.

The relaxation times as calculated are quite sensitive to molecular weight. Since the theory assumes a monodisperse polymer, some divergence of theory from experimental data would be expected even for fractionated polymers. A more fundamental cause for possible disagreement is the fact that the theory does not take specific account of the interactions between remotely connected parts of a polymer chain. These interactions make it impossible for the chain to assume many of the "random flight" configurations[10] and, in addition, make necessary a temporary cooperation between the motions of segments which happen to come into contact.

MEASUREMENT

A plane shearing wave is generated in a liquid by the oscillatory motion in its own plane of a plane surface in contact with the liquid. If the motion is sinusoidal and has the angular frequency $\omega=2\pi f$, the velocity of the surface can be represented by

$$v_s=v_oe^{i\omega t}. \tag{4}$$

The transverse velocity of the liquid at the distance x from the oscillating surface is

$$v_l=v_oe^{-\Gamma x}e^{i\omega t}, \tag{5}$$

The propagation constant Γ is

$$\Gamma=(i\omega\rho/\eta^*)^{\frac{1}{2}},$$

where i is $(-1)^{\frac{1}{2}}$, ρ is the density of the liquid and η^* is its complex viscosity. The force exerted by a unit area of the surface upon the liquid is

$$F=-\eta^*(\partial v_l/\partial x)_{x=0}=\eta^*\Gamma v_s.$$

The complex mechanical impedance Z_M which the liquid presents to the shearing force is given by the ratio of the stress to the particle velocity,

$$Z_M=R_M+iX_M=(i\omega\rho\eta^*)^{\frac{1}{2}}. \tag{6}$$

The components of the complex viscosity and complex rigidity are related to the components, R_M and X_M, of the mechanical impedance by the equations

$$\eta_1=\frac{G_2}{\omega}=\frac{2R_MX_M}{\omega\rho}, \tag{7a}$$

$$\eta_2=\frac{G_1}{\omega}=\frac{R_M^2-X_M^2}{\omega\rho}, \tag{7b}$$

For a normal liquid $\eta_2=0$ and

$$R_M=X_M=(\pi f\rho\eta)^{\frac{1}{2}}.$$

The techniques used in this work make use of torsionally oscillating bodies which present cylindrical surfaces to the liquid under measurement. The tangential motions of the surfaces generate shearing waves in the liquids. Description of these waves by Eq. (5), which applies strictly only to plane waves, is permissible because the waves are damped out in distances which are very small compared to the radii of curvature of the surfaces. The properties of the liquid are determined by

[10] P. J. Flory, J. Chem. Phys. **17**, 303 (1949); J. J. Hermans, Rec. trav. chim. **69**, 220 (1950); E. W. Montroll, J. Chem. Phys. **18**, 734 (1950); P. Debye and R. J. Rubin, Phys. Rev. **87**, 214 (1952); R. J. Rubin, J. Chem. Phys. **20**, 1940 (1952).

measurements of the energy dissipated in the shearing wave and of the change of the resonant frequency of the oscillator caused by the liquid. From the energy dissipation the real part R_M of the impedance of the liquid is calculated, and from the frequency shift caused by the mass loading the imaginary part X_M is obtained.

The data presented here were obtained by use of two techniques. At the frequencies of 220 cps, 900 cps, and 2300 cps, the torsion pendulum, which is described fully in another publication,[3] was used. Measurements at 20 kilocycles and 60 kilocycles were obtained by use of the torsion crystal technique, which was introduced by W. P. Mason.[2]

The shearing gradients generated in the liquids by these oscillators are fairly small. With the torsion crystals the maximum gradients generated are of the order of one sec^{-1}. The maximum angular displacement of the torsion pendulums is estimated to be about 10^{-4} radian. The maximum shearing gradients generated by oscillations of that amplitude in a liquid of 0.5-centipoise viscosity are calculated to be 150 sec^{-1} at 220 cps, 1000 sec^{-1} at 900 cps, and 4000 sec^{-1} at 2300 cps.

TORSION PENDULUM[3]

The torsion pendulums used in this work consisted of a hollow lower disk, 4 cm in diameter, and a massive upper disk, 14 cm in diameter, connected by a torsion rod, the diameter of which was machined to give the desired frequency. The pendulum, which was made of stainless steel, was suspended from the top of a chamber by a steel wire 0.010 in. in diameter. The temperature of the pendulum was controlled by circulation of water from a thermostatic bath through the walls of the chamber. The liquid to be measured was contained in the hollow disk. The torsional oscillations of the hollow disk were excited by the action of an alternating magnetic field upon an Alnico V magnet embedded in the lower disk above the chamber which held the sample. The upper disk which had a moment of inertia 700 times as great as that of the lower disk, served as a reference against which the oscillations of the lower disk were measured. Mirrors were mounted on both disks and an optical detecting system was arranged in such a way that the light striking a multiplier phototube was modulated by the oscillations of the pendulum. The excitation was discontinued and the oscillations were allowed to decay. The alternating voltage which represented the decaying oscillations of the pendulum was applied to a system of electronic voltage discriminators which started and stopped a counter and a chronograph at predetermined voltage levels corresponding to certain amplitudes of the oscillations. The counter registered the integral number of oscillations which occurred between the two voltage levels and the chronograph registered the time required for that number of complete oscillations with an uncertainty of $\pm 0.625\times 10^{-6}$ sec. From these data the angular frequency ω and damping constant α of the free oscillations of the pendulum were calculated.

The values of R_M and X_M for a polymer solution were calculated from the equations

$$R_M=(\pi f\rho_s\eta_s)^{\frac{1}{2}}+\frac{\alpha_p-\alpha_s}{K_\alpha}, \tag{8a}$$

$$X_M=(\pi f\rho_s\eta_s)^{\frac{1}{2}}+\frac{\omega_s-\omega_p}{K_\omega}, \tag{8b}$$

where the subscripts s and p refer, respectively, to the solvent and the solution. Values of the constants K_α and K_ω were determined for each pendulum by empirical calibration with standard liquids whose viscosities covered the range from 0.5 to 10 centipoise.

TORSION CRYSTAL[2,11]

The torsion crystals used in this work were cylinders of quartz 10 cm long and 1 cm in diameter cut with the axis of the cylinder parallel to the x axis of the quartz. These crystals were used at their fundamental frequency of 19 600 cps and at their third harmonic of 58 800 cps. The crystals were supported by four phosphor bronze wires 0.006 in. in diameter. These wires were soldered to spots of silver paint which were spaced 90 degrees apart around the cylinder midway between the ends. Each wire formed the electrical connection to one of four quadrantal electrodes of evaporated gold which ran the length of the crystal.

In use the crystals were immersed in the liquid in a closed stainless steel tube. The tube was supported in a thermostatic water bath. The crystal was connected electrically as the unknown arm of a precision capacitance-conductance bridge. The effective electrical resistance R of the crystal at series resonance and its resonant frequency f_r were measured. The frequency was determined with an electronic counter and chronograph system which was precise and accurate to approximately one part in 5×10^6. The two frequencies at which the resistance of the crystal was twice its resistance at resonance were also determined. The mean of these frequencies served as a check on the measured value of the resonant frequency and the difference between them gave the band width $(\Delta f)_{\frac{1}{2}}$ of the resonance curve of the crystal.

The value of R_M for each polymer solution was calculated from the equations[11]

$$R_M=(\pi f\rho_s\eta_s)^{\frac{1}{2}}+\frac{R_p-R_s}{K_1}, \tag{9}$$

$$K_1=\frac{4L_c}{\rho_c}\left[\frac{1}{r}+\frac{1}{l}\right], \tag{9a}$$

where R_s, R_p = electrical resistance of crystal in solvent

[11] Rouse, Bailey, and Sittel (to be published).

and solution, respectively, and r, l, ρ_c = radius, length, and density of the crystal, respectively. L_c, the equivalent electrical inductance of the crystal, is related to the electromechanical coupling constant of the crystal. A value of L_c was calculated from each set of measurements on each liquid by means of the equation[11]

$$L_c = R/2\pi(\Delta f)_{\frac{1}{2}},$$

where R is the equivalent electrical resistance of the crystal, measured at series resonance. The value of L_c was a constant for each crystal at each harmonic and was independent of the liquid in which the crystal was immersed.

The value of X_M for each polymer solution was calculated from the equations[11]

$$X_M = (\pi f \rho_s \eta_s)^{\frac{1}{2}} + \frac{f_s - f_p}{K_2} + C, \tag{10}$$

$$K_2 = \frac{1}{\pi \rho_c}\left[\frac{1}{r} + \frac{1}{l}\right], \tag{10a}$$

where f_s and f_p are the resonant frequencies of the crystal in solvent and solution, respectively. C is an empirical correction term made necessary by the roughness of the surface of the crystal.[11] This correction was ignored in calculation of a first approximation to X_M. The value of C was then calculated from the equation

$$C = A(\rho_s - \rho_p) + B\left[\frac{X_M \omega \rho_p^2}{R_M^2 + X_M^2} - \left(\frac{\omega \rho_s^3}{2\eta_s}\right)^{\frac{1}{2}}\right]. \tag{10b}$$

The values of the constants A and B were determined empirically for each crystal at each harmonic by fitting an equation of the form

$$f = f_0 - K_2\left\{(\pi f \rho \eta)^{\frac{1}{2}} + \rho\left[A - B\left(\frac{\omega \rho}{2\eta}\right)^{\frac{1}{2}}\right]\right\}, \tag{11}$$

to data obtained for a series of pure liquids covering a range of viscosities and densities.

MATERIALS

The polymers used in this work were fractions of polystyrene and of polyisobutylene prepared by the procedure[12] of adding nonsolvent to produce a slight, permanent turbidity, redissolving the polymer by increasing the temperature, and precipitating the fraction by slow cooling to the original temperature. The solvent-precipitant system used for polystyrene was methyl ethyl ketone-methyl alcohol and for the polyisobutylene was benzene-acetone. The fractionations were carried out at 30°C.

Four polystyrenes were used in the fractionation. One, which was prepared by emulsion polymerization at −10°C with ferrous pyrophosphate activator and benzoyl peroxide catalyst, had a viscosity average molecular weight[12] of 4 000 000. In a single precipitation fractionation of this polymer from a solution with an initial concentration of 3 grams/liter seven fractions of appreciable size were obtained. Their viscosity average molecular weights ranged from 6 300 000 to 2 600 000. The second fraction, which had a molecular weight of 6 200 000 was used in this work.

The other polystyrenes used in the fractionations were prepared by emulsion polymerization at 45°C with potassium persulfate catalyst and primary dodecyl mercaptan as modifier. The viscosity average molecular weights were 105 000, 210 000 and 570 000. These polymers were fractionated from solutions with original concentrations of approximately 1 gram/liter. Refractionations of these fractions were then carried out from solutions, the original concentrations of which varied from 0.7 gram/liter to 0.1 gram/liter with increasing molecular weight. Two of these fractions were used in the work reported here.

The fractions of polyisobutylene were obtained by a single fractionation of a polymer (supplied by the Enjay Company) known commercially as "Vistanex *LM-MS*," which had a viscosity average molecular weight of 63 000. The initial concentration of the solution was 3 grams/liter. Fractions with viscosity average molecular weights ranging from 10 000 to 326 000 were obtained. Those used in this work had molecular weights of 127 000 to 165 000.

The molecular weights of the polystrenes were calculated from their intrinsic viscosities $[\eta]$ in toluene using the equation[13]

$$\log \bar{M}_v = (\log[\eta] + 4.01)/0.73,$$

where $\bar{M}_v$ is the viscosity average molecular weight.[12] Molecular weights of the polyisobutylenes were calculated from their intrinsic viscosities in carbon tetrachloride using the equation[14]

$$\log \bar{M}_v = (\log[\eta] + 3.345)/0.64.$$

The viscosities were measured at 30.35±0.05°C using Cannon-Fenske-Ostwald viscometers which had been calibrated by the National Bureau of Standards.

DISCUSSION OF RESULTS

The results of the measurements on solutions of these polymers are presented in terms of the real parts of the complex viscosity and the complex shear modulus of the solutions. The curves in the figures were calculated by means of Eqs. (1), (2), and (3) from the steady-flow viscosities of the solution and solvent, the molecular weight and concentration of the polymer, and the absolute temperature. The behavior of the experimental data

[12] P. J. Flory, J. Am. Chem. Soc. **65**, 372 (1943).

[13] T. G. Fox, Jr., and P. J. Flory, J. Am. Chem. Soc. **73**, 1915 (1951).

[14] T. G. Fox, Jr., and P. J. Flory, J. Am. Chem. Soc. **70**, 2384 (1948).

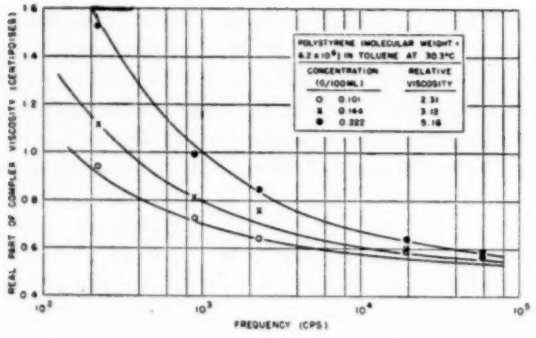


FIG. 1. Effect of concentration on real part of complex viscosity.

conforms fairly well to the quantitative predictions of the theory.

The precision and accuracy of the data presented here are believed to be somewhat less than will be obtained with these techniques after the addition of certain refinements. Limitation of the time available for this work made it necessary to carry out these measurements without these refinements. The results can, therefore, be considered to provide only a preliminary test of the theory. However, within this limitation the data indicate that the predictions of the theory are a good first approximation to the actual response of solutions of linear polymers to oscillatory shearing stresses.

In addition to providing a preliminary confirmation of the theory, the data serve to illustrate several aspects of the relation of the viscoelastic properties of solutions of polymers to the properties of the polymer and the solvent. The effects of the concentration and molecular weight of the polymer are discussed below with reference to the data for solutions of various polystyrene fractions in toluene; the relation of the viscoelastic processes to the viscosity and solvent power of the solvent are demonstrated by the data obtained for polyisobutylene fractions in several solvents.

In discussing these relationships it is helpful to consider the relaxation times as represented not only by Eq. (3) but also by the following equation from which

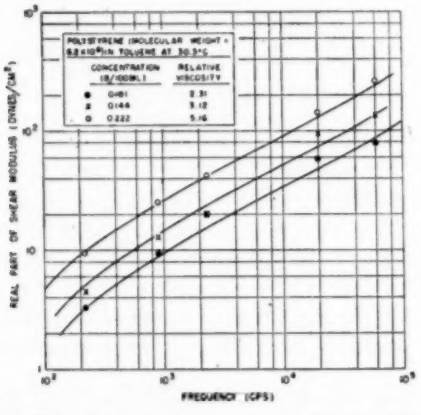


FIG. 2. Effect of concentration on real part of complex shear modulus.

Eq. (3) was derived:[7]

$$\tau_p = NS^2/6\pi^2 p^2 BkT. \tag{12}$$

In this expression N is, as before, the number of submolecules per molecule, S^2 is the mean square separation of the ends of the chain, and B is the average mobility (velocity under unit force) of the ends of the submolecules.

The effects of increasing the concentration of polymer upon the dynamic properties of solutions of a polymer are shown in Figs. 1 and 2 for toluene solutions of a polystyrene fraction with a molecular weight of 6 200 000. One effect was the increase by a factor of 1.44 of the relaxation times as calculated from Eq. (3); it was caused by the fact that the quantity $(\eta_o - \eta_s)/n$ increased with concentration. Equation (12) suggests that this increase in the relaxation times was caused by a decrease in the mobility B, resulting from temporary interferences between the parts of different polymer molecules.

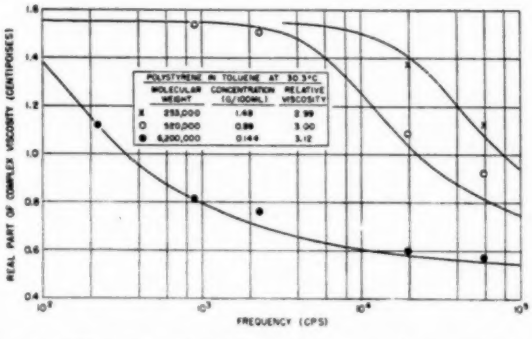


FIG. 3. Effect of molecular weight on real part of complex viscosity of solutions having (approximately) equal viscosities in steady flow.

The effect of the increase in relaxation times was to displace the dispersion curves for η_1 and G_1 to lower frequencies. This accounts for a portion of the increase in the real part of the rigidity, since at the same frequency a larger fraction of the contribution of each relaxation process was effective. A larger portion of the increase in G_1 was the result merely of the larger number of molecules per unit volume of solution; according to Eq. (2) the contribution of each relaxation process to the modulus is given by nkT.

The effects of increasing the molecular weight of the polymer while holding the steady flow viscosity of the solution constant are shown in Figs. 3 and 4. The solutions, which contained polystyrene fractions in toluene, each had a relative viscosity of approximately 3. The number of molecules per unit volume of solution for polymers having molecular weights of 6 200 000, 520 000 and 253 000 were in the ratio 1:70:250. This resulted in a wide variation of the calculated relaxation times which, for this series of solutions, were approximately inversely proportional to n. The displacement of the dispersion curves along the frequency scale by the

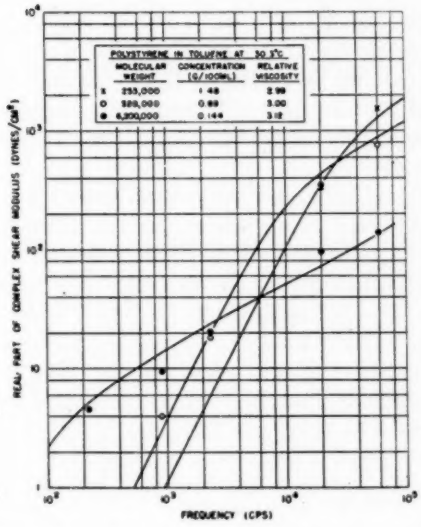


FIG. 4. Effect of molecular weight on real part of complex shear modulus of solutions having (approximately) equal viscosities in steady flow.

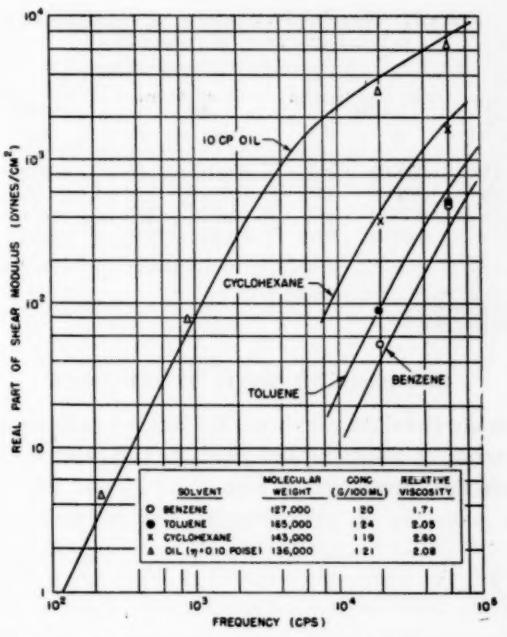


FIG. 5. Effect of solvent on real part of complex shear modulus of solutions of polyisobutylene at 30.3°C.

differences in the relaxation times is shown clearly in Fig. 3.

The frequency at which the modulus G_1 first became measurable decreased with increasing relaxation times, as shown in Fig. 4. Because of the differences in the numbers of molecules in these solutions, the contribution of each relaxation process to the modulus increased with decreasing molecular weight. This caused the moduli of the solutions of the lower-molecular-weight polymers to rise above the modulus of the solution of the high-molecular-weight polymer with increasing frequency.

The effects of the solvent upon the viscoelastic properties of solutions of polyisobutylene are shown in Fig. 5 and Table I. The solutions contained 12 grams/liter of polyisobutylene fractions with molecular weights between 127 000 and 165 000. The relaxation times calculated from Eq. (3) account satisfactorily for differences in the locations of the dispersion curves along the frequency axis. The differences in the relaxation times arose mostly from differences in the values of $(\eta_o - \eta_s)$.

The lateral displacements of the curves in Fig. 5 with respect to one another illustrate the differences in the relaxation times in the different solutions. The physical causes of these differences are suggested by Eq. (12). The two factors in that equation which reflect the differences produced by change of solvent are B, the mobility of the end of a submolecule, and S^2, the mean-square separation of the ends of the chains. Relative values of S^2 were estimated from the intrinsic viscosities of the polymers in the various solvents on the basis of Flory's conclusion[10] that S^2 is proportional to $[\eta]^{\frac{2}{3}}$. From these values and the relaxation times calculated from Eq. (3) relative values of the mobilities in the various solvents were estimated by means of Eq. (12). These estimated mobilities were found to be roughly inversely proportional to the viscosities of the corresponding solvents. These considerations lead to the following conclusions. (1) Of two solvents with the same viscosity the one in which S^2 has the larger value will give solutions having longer relaxation times. (2) Of two solvents with the same thermodynamic solvent properties the more viscous will give a lower value of B and, consequently, longer relaxation times. These conclusions will not apply at frequencies so high that the rates of internal rotation about chemical bonds become important.

TABLE I. Effect of solvent on real part of complex viscosity of solutions of polyisobutylene. 30.3°C.

Solvent	Molecular weight	Concentration g/100 ml	Relative viscosity	Real part of complex viscosity[a] 220 cps Meas.	220 cps Theory	900 cps Meas.	900 cps Theory	19.6 kc Meas.	19.6 kc Theory	58.8 kc Meas.	58.8 kc Theory
Cyclohexane	143 000	1.19	2.60	...	...	...	...	0.0197	0.0201	0.0169	0.0159
Benzene	127 000	1.20	1.71	...	...	...	...	0.0095	0.0095	0.0091	0.0092
Toluene	165 000	1.24	2.05	...	...	...	...	0.0099	0.0105	0.0093	0.0097
Oil[b]	136 000	1.21	2.08	0.202	0.207	0.196	0.205	0.147	0.138	0.134	0.122

[a] Poises.
[b] Oil had viscosity of 0.099 poise at 30.3°C.

TABLE II. Viscoelastic properties[a] of unfractionated polyisobutylene at 20 kc and 25°C.

Concentration g/100 ml	From experimental points[b] η_1	G_1	From smooth curves[b] η_1	G_1	From theoretical equations[c] η_1	G_1
0.125	0.0092	160	0.0097	94	0.0096	65
0.250	0.0099	170	0.0104	200	0.0106	160
0.475	0.0111	450	0.0118	400	0.0121	330
1.00	0.0160	910	0.0152	970	0.0169	860

[a] η_1 in poises; G_1 in dynes/cm².
[b] See reference 4.
[c] See reference 7.

COMPARISON WITH OTHER DATA

Among the data presented by Baker,[4] only those for one series of solutions are given in sufficient detail to permit the application of Eqs. (1), (2), and (3). These were solutions in cyclohexane of a polyisobutylene which had an $\bar{M}_v$ of 1.18×10^6; the measurements reported were made at 20 kc and 25°C. Values of η_1 and G_1 were calculated by means of Eqs. (7a) and (7b), using values of R_M and X_M estimated from the graph given by Baker.[4] The results of these calculations are shown in Table II.

A third set of values for η_1 and G_1 for these solutions were calculated by means of the theoretical Eqs. (1), (2), and (3), using the data of Baker's paper[4] to estimate the steady-flow viscosities. The results of these calculations are also given in Table II. They differ from the values calculated from the experimental data by little more than the apparent experimental error.

The work discussed herein was performed as a part of the research project sponsored by the Reconstruction Finance Corporation, Office of Synthetic Rubber, in connection with the Government Synthetic Rubber Program.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Temperature Coefficients of Non-Newtonian Viscosity at Fixed Shearing Stress and at Fixed Rate of Shear*†

A. B. BESTUL AND H. V. BELCHER
National Bureau of Standards, Washington 25, D. C.

(Received March 23, 1953)

Viscosity decreases with increasing shearing stress or rate of shear in the non-Newtonian flow of high polymeric systems. It is shown that it follows from this fact that the variation of viscosity with temperature is greater at fixed shearing stress than at fixed rate of shear.

For such systems, it is shown experimentally that the variation of viscosity with temperature at fixed shearing stress is independent of shearing stress. However, the variation at fixed rate of shear is always less than that at very low shearing stress, and decreases with increasing rate of shear.

This behavior is reproduced by Eyring's hyperbolic sine flow equation, which is the antecedent for most other, more adequate, non-Newtonian flow equations. However, this equation is shown to be inadequate in that it predicts incorrect values of shearing stress for the occurrence of non-Newtonian behavior.

I. INTRODUCTION

THE study of the effect of variation of temperature on the viscous behavior of non-Newtonian materials may be expected to lead to important information about the molecular structure of these materials. In order to establish a procedure for such a study, this paper defines two partial derivatives which are convenient to examine. Qualitative and quantitative experimental data are used to investigate the behavior of these derivatives for high polymeric systems. The results of this investigation are applied as an experimentally established criterion to test the adequacy of Eyring's hyperbolic sine flow equation with respect to the effect of temperature variation. This equation is tested as an antecedent of more adequate equations, and not for its own sake.

The expectation of such a study leading to important information about molecular structure is supported by the fact that similar studies for Newtonian materials has led to such information. For a Newtonian material, viscosity (η) is independent of shearing stress (s) or rate of shear ($\dot{\gamma}$), and is a constant at any one temperature. Therefore, for such a material, the variation of viscosity with temperature (T) is singly defined and is obtained without regard to any special conditions of shearing stress or rate of shear. This variation is often expressed satisfactorily by the Arrhenius equation

$$\eta = A\exp(-\Delta E/RT), \tag{1}$$

where R is the gas constant, and A and ΔE can be considered as constants over limited temperature ranges. The quantity ΔE can be viewed as the energy of activation for flow, and A depends largely on the

* This work was performed as part of a research project sponsored by the Reconstruction Finance Corporation, Office of Synthetic Rubber, in connection with the Government Synthetic Rubber Program.

† Presented at the 1952 Annual Meeting of the Society of Rheology, Philadelphia, Pennsylvania, October 30–31, 1952.

entropy of activation for flow. Since viscosity is defined as the ratio of shearing stress to rate of shear, the relation between these latter two quantities is expressed by the proportionality

$$s=(A\exp(-\Delta E/RT))\dot{\gamma}. \tag{2}$$

For a non-Newtonian material, however, viscosity at fixed temperature is dependent on shearing stress or rate of shear. If it is desired to retain the form of Eq. (1), either A, or ΔE, or both must be considered as functions of shearing stress or rate of shear. Accordingly, viscosity is no longer a single valued function of temperature. In order to make any quantitative evaluation of the variation of viscosity with temperature, it is therefore necessary to deal with variations under some kind of fixed conditions of shearing stress and rate of shear. Several different specifications of these fixed conditions may be made. This paper examines the two simplest specifications, namely, those of fixed shearing stress and of fixed rate of shear. The specification of either one of these parameters is sufficient to specify both of them completely.

The applicability of Eq. (1) for Newtonian materials makes it convenient to deal with the derivative $d\ln\eta/d(1/T)$ for such materials. In order to compare results for non-Newtonian flow at fixed s and at fixed $\dot{\gamma}$ with results for Newtonian flow, it is sometimes convenient to use the derivatives $[\partial\ln\eta/\partial(1/T)]_s$ and $[\partial\ln\eta/\partial(1/T)]_{\dot{\gamma}}$ for non-Newtonian flow. This usage amounts to arbitrarily throwing the whole variation of η with s or $\dot{\gamma}$ into ΔE.

II. CONSIDERATION OF QUALITATIVE INFORMATION

In the type of non-Newtonian viscous behavior exhibited by high polymers and their solutions, the variation of viscosity with temperature at fixed shearing stress is higher than that at fixed rate of shear. This conclusion can be reached either graphically or analytically.

The above-mentioned relation can readily be seen from Fig. 1. This figure is a typical family of logarithmic plots of rate of shear *versus* shearing stress for a high polymeric system at several different temperatures. The variation at fixed shearing stress is obtained by a vertical cross plot, whereas the variation at fixed rate of shear is obtained by a horizontal cross plot. It is clear that the distances between the different temperatures are greater along the vertical cross plots than along the horizontal cross plots. This relation results because the slopes of the plots are greater than one and the plots for different temperatures tend to lie very roughly parallel to each other.

This same conclusion, and something more, can be reached analytically. Formal differentiation of the defining relation for viscosity with respect to temperature at fixed shearing stress and at fixed rate of shear gives expressions for the variation at fixed shearing stress and that at fixed rate of shear. The ratio of these

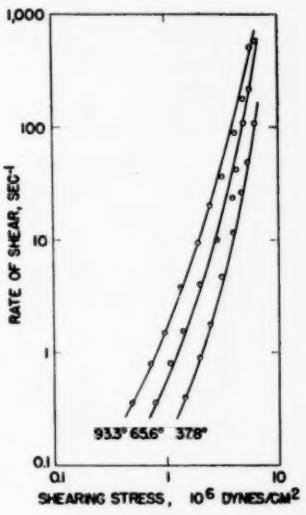


FIG. 1. Rate of shear *versus* shearing stress for X-672 GR-S rubber at three different temperatures.

variations is found to be always greater than one whenever viscosity decreases with increasing shearing stress or rate of shear, as is the case for high polymeric systems.

The differentiation proceeds either by considering shearing stress as a function of rate of shear and temperature, or by considering rate of shear as a function of shearing stress and temperature. This procedure involves a purely artificial separation, but it is entirely permissible and is used here to simplify the analysis.

The first case above can be expressed as

$$\eta=s(\dot{\gamma}, T)/\dot{\gamma}=\eta(\dot{\gamma}, T), \tag{3}$$

and the second as

$$\eta=s/\dot{\gamma}(s, T)=\eta(s, T). \tag{4}$$

By formally differentiating Eq. (3) with respect to temperature at fixed shearing stress it is possible to obtain the relation

$$(\partial\eta/\partial T)_{\dot{\gamma}}/(\partial\eta/\partial T)_s=(\partial\ln\eta/\partial\ln\dot{\gamma})_T+1. \tag{5}$$

Also, by formally differentiating Eq. (4) with respect to temperature at fixed rate of shear it is possible to obtain the relation

$$(\partial\eta/\partial T)_s/(\partial\eta/\partial T)_{\dot{\gamma}}=1-\dot{\gamma}(\partial\eta/\partial s)_T. \tag{6}$$

The consistency of Eqs. (5) and (6) can be demonstrated by equating the product of their right-hand sides to unity, which is certainly the product of their left-hand sides since they are reciprocals. From this equation it is easy to arrive at the identity

$$1-(\partial\ln\dot{\gamma}/\partial\ln s)_T\equiv 1-(\partial\ln\dot{\gamma}/\partial\ln s)_T. \tag{7}$$

From Eq. (6) it follows that for materials of the type with which we are concerned $(\partial\eta/\partial T)_s/(\partial\eta/\partial T)_{\dot{\gamma}}$ must always be greater than or equal to one, since $(\partial\eta/\partial s)_T$ is always negative or zero, i.e., viscosity either decreases or remains stationary as shearing stress increases. Independently, from Eq. (5) it follows that

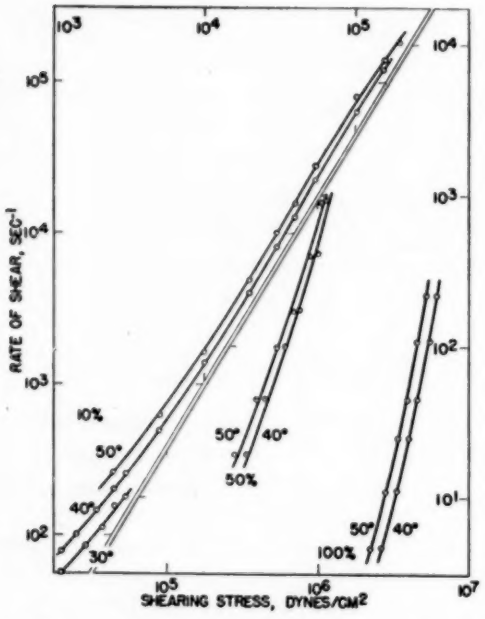


FIG. 2. Rate of shear *versus* shearing stress for solutions of X-518 GR-S rubber in alpha-methylnapthalene at three concentrations and three temperatures.

for these materials $(\partial\eta/\partial T)_{\dot\gamma}/(\partial\eta/\partial T)_s$ must always be less than or equal to one since $(\partial\ln\eta/\partial\ln\dot\gamma)_T$ is always negative or zero, i.e., viscosity either decreases or remains stationary as rate of shear increases. Furthermore, $(\partial\eta/\partial T)_{\dot\gamma}/(\partial\eta/\partial T)_s$ is never negative since $(\partial\ln\eta/\partial\ln\dot\gamma)_T$ cannot be more negative than minus one, as seen from the relation

$$(\partial\ln s/\partial\ln\dot\gamma)_T=(\partial\ln\eta/\partial\ln\dot\gamma)_T+1, \qquad (8)$$

which follows from either Eq. (3) or (4). From Eq. (8), if $(\partial\ln\eta/\partial\ln\dot\gamma)_T$ were more negative than minus one, then $(\partial\ln s/\partial\ln\dot\gamma)_T$ would be negative, i.e., rate of shear would decrease with increasing shearing stress. Such a behavior seems unlikely enough to be rejected, so $(\partial\eta/\partial T)_{\dot\gamma}/(\partial\eta/\partial T)_s$ can be considered as never negative.

The condition that the ratio of the two types of temperature variations concerned must never be negative is consistent with the fact that both of the variations themselves are usually negative. Moreover, this condition also requires that if either of these types of variations should change sign and become positive under any circumstances, it would be necessary for the other type of variation to also change sign under the same circumstances. The likelihood that either type of variation should ever change sign and become positive does not seem very large for high polymeric systems, but such a possibility is conceivable, and the above requirement is of interest in connection with a point to be discussed later.

It is possible therefore to make the perfectly general conclusion that in the non-Newtonian viscous behavior of high polymeric systems the temperature variation of viscosity at fixed shearing stress is greater than this variation at fixed rate of shear. This conclusion, however, says nothing about the absolute magnitude of either of these two types of variations. It does not seem possible to make any firm theoretical predictions about the absolute magnitude of these variations. It is necessary therefore to turn to an examination of experimental data to obtain information about the magnitudes of the variations.

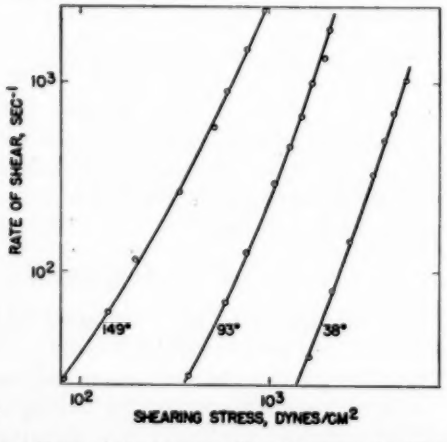


FIG. 3. Rate of shear *versus* shearing stress for Vistanex LM-S polyisobutylene at three different temperatures

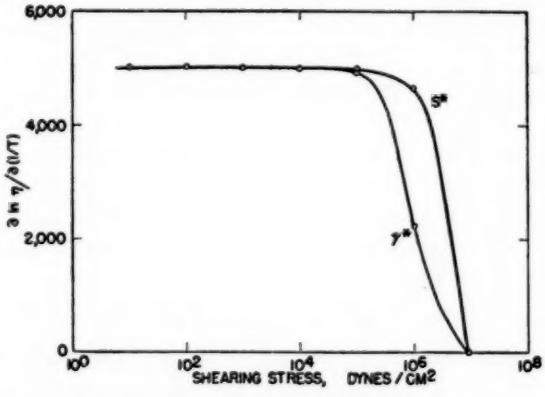


FIG. 4. Temperature variation of viscosity for hypothetical substance at fixed shearing stress (s^*) and at fixed rate of shear (γ^*) according to hyperbolic sine flow equation.

III. CONSIDERATION OF QUANTITATIVE INFORMATION

Experimental data are presented here on the variation of viscosity with temperature at fixed shearing stress and at fixed rate of shear for the following three high polymeric systems:

(1) Undiluted Vistanex LM-S polyisobutylene (viscosity-average molecular weight about 70 000).

(2) Solutions of 10 percent, 50 percent and 100 percent (undiluted polymer) X-518 GR-S (24 percent styrene, 76 percent butadiene) rubber (viscosity-average molecular weight about 200 000) in alpha-methylnaphthalene.

(3) Undiluted X-672 GR-S (21 percent styrene, 79 percent butadiene) rubber (viscosity-average molecular weight about 200 000).

The data for systems (1) and (2) were obtained previously as incidental parts of two investigations directed primarily at different topics than the present one.[1,2]

The measurements for all three systems were made with the McKee Consistometer.[3,4] This instrument is a short capillary viscometer in which a piston-cylinder arrangement is used to force the experimental material through a capillary or several capillaries in parallel. The details of the instrument and its operation are given in the above-listed references. The measurements were reduced to rate of shear and shearing stress by the use of the differentiation method,[5] for which the working equation is

$$\dot{\gamma}_w = (D/4)(3 + d\log D/d\log s_w), \qquad (9)$$

where $\dot{\gamma}_w$ is the rate of shear at the capillary wall, and D and s_w are the nominal rate of shear and the shearing stress at the capillary wall as calculated from the usual relations of Poiseuille.

TABLE I. Apparent flow activation energies for Vistanex LM-S polyisobutylene

Condition	E^*, kcal
Literature value, E	12.5
Extrapolated to $s=0$, E_0	11
At fixed $s=800\,000$ dynes/cm², E^*_s	12.3
At fixed $\dot{\gamma}=200$ sec^{-1}, $E^*_{\dot{\gamma}}$	5.9
At fixed $\dot{\gamma}=2000$ sec^{-1}, $E^*_{\dot{\gamma}}$	3.5

The data for these systems are shown in Figs. 1, 2, and 3, respectively.

Tables I, II, and III, respectively, list the variations of viscosity with temperature for these systems at various conditions of fixed shearing stress and of fixed rate of shear. The variations are expressed as apparent flow activation energies, E^* given by $R[\partial \ln \eta/\partial(1/T)]$. These quantities are used only as a convenient and readily compared means of expressing the variations, and their use is not necessarily intended to imply that in non-Newtonian flow they retain the full significance attached to activation energies in Newtonian flow. Log viscosity is plotted *versus* reciprocal temperature at fixed s and fixed $\dot{\gamma}$, and the values of E^* are $2.3R$ times the slopes of these plots. These values have the advantage that they can be directly compared with the

TABLE II. Apparent flow activation energies for X-518 GR-S rubber solutions in alpha-methylnaphthalene.

Concentration, wt percent	E_0 kcal	E^*_s, kcal	s, dynes/cm²	$E^*_{\dot{\gamma}}$, kcal	$\dot{\gamma}$, sec^{-1}
100	13.0				
		11.1	2 500 000		
		13.9	4 600 000		
				3.7	6.3
				2.8	160
50	9.2				
		8.8	320 000		
		10.6	1 100 000		
				4.2	20
				2.3	630
10	4.1				
		4.2	2500		
		3.7	20 000		
		3.7	160 000		
				3.7	250
				2.3	4000
				2.3	126 000

variations of viscosity with temperature for the systems concerned at extremely low shearing stresses. This advantage arises because the variations at low shearing stresses are usually expressed as flow activation energies since Newtonian conditions are approximated at low shearing stresses. When the slopes are obtained from two temperatures only, nothing is learned about the linearity of the $\log \eta$ *versus* $1/T$ plot. However, the data shown in Fig. 1 for the X-672 GR-S rubber cover a conveniently large range of rate of shear and shearing stress at three different temperatures so these are examined in terms of linearity of such plots as is mentioned later.

The variations of viscosity listed in Tables I, II, and III were obtained from Figs. 1, 2, and 3 by the use of vertical and horizontal cross plots as mentioned in the Sec. I.

The apparent flow activation energies for Vistanex LM-S are listed in Table I. Around the temperatures used here, the value at an extremely low shearing

TABLE III. Viscosities of X-672 GR-S rubber at fixed shearing stresses.

s, dyne/cm² $\times 10^{-6}$	1/°K $\times 10^3$	$\log \eta$ (poises)	$\Delta \log \eta/\Delta(1/°K)$ $\times 10^3$	E^*_s, kcal
1.0	2.73	5.84		
			1.77	8.1
	2.95	6.23		
1.5	2.73	5.57		
			1.95	
	2.95	6.00		9.7
			2.26	
	3.22	6.61		
3.0	2.73	4.95		
			1.91	
	2.95	5.37		8.9
			1.96	
	3.22	5.90		
5.0	2.73	4.24		
			1.32	
	2.95	4.53		8.7
			2.37	
	3.22	5.17		

[1] A. B. Bestul and H. V. Belcher, J. Research Natl. Bur. Standards (to be published).

[2] Bestul, Belcher, Quinn, and Bryant, J. Phys. Chem. **56**, 432 (1952).

[3] S. A. McKee and H. S. White, Am. Soc. Testing Materials Bull. No. 153, 90 (1948).

[4] S. A. McKee and H. S. White, J. Research Natl. Bur. Standards **46**, 18 (1951).

[5] See, e.g., T. Alfrey, Jr., *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948), p. 41.

TABLE IV. Viscosities of X-672 GR-S rubber at fixed rates of shear.

$\dot{\gamma}$, sec^{-1} ×10^{-6}	1/°K ×10^3	log η (poises)	$\Delta\log\eta/\Delta(1/°K)$ ×10^3	$E^*_{\dot{\gamma}}$, kcal
0.3	2.73	6.16		
			1.00	
	2.95	6.38		4.8
			1.07	
	3.22	6.67		
1.0	2.73	5.92		
			0.86	
	2.95	6.11		3.7
			0.74	
	3.22	6.31		
5.0	2.73	5.51		
			0.68	
	2.95	5.66		2.8
			0.56	
	3.22	5.81		
10.0	2.73	5.32		
			0.59	
	2.95	5.45		2.4
			0.48	
	3.22	5.58		
50.0	2.73	4.82		
			0.45	
	2.95	4.92		2.0
			0.41	
	3.22	5.03		
100.0	2.73	4.60		
			0.32	
	2.95	4.67		1.8
			0.44	
	3.22	4.79		

stress, is 12.5 kcal.[6] The value of 11 kcal for the present data extrapolated to zero shearing stress agrees roughly with the above low shear value. This extrapolation was made on the basis of the relation

$$1/\eta = a + bs^2, \tag{10}$$

where a and b are constants, a being the reciprocal of the viscosity at zero shearing stress. The applicability of this relation to the data concerned was verified by the linearity of plots of reciprocal viscosity *versus* shearing stress squared. The reciprocal viscosities at zero shearing stress were obtained as the intercepts of these plots on the reciprocal viscosity axis. The apparent flow activation energy, E^*_s at a fixed shearing stress of 800 000 dynes/cm^2 is 12.3 kcal, which is about the same as the above value from literature[6] and that for the present data extrapolated to zero shearing stress. However, under fixed rate of shear the apparent flow activation energies, $E^*_{\dot{\gamma}}$, are 5.9 kcal at 200 sec^{-1} and 3.5 kcal at 2000 sec^{-1}. These values are both smaller than the low shearing stress value, and they show a decrease as one increases the rate of shear at which they are calculated.

Apparent flow activation energies for the solutions of X-518 GR-S rubber are listed in Table II. The values for low shearing stresses, E_0 are derived from results of Cragg, Faichney, and Olds.[7] At each concentration, the values for E_s^* suggest that E_s^* changes but little with a change in the value of shearing stress at which it is calculated, and that at any value of shearing stress E_s^* is about the same as E_0. The values shown for $E^*_{\dot{\gamma}}$ support the idea that $E^*_{\dot{\gamma}}$ is always smaller than E_0, and that $E^*_{\dot{\gamma}}$ decreases as one increases the rate of shear at which it is calculated.

The data for the X-672 GR-S rubber, shown in Fig. 1 and in Tables III and IV, were obtained over as wide as possible a range of the experimental variables in order to provide as thoroughly as possible an examination of the behavior. Tables III and IV give values of log viscosity as functions of reciprocal temperature at several fixed values of shearing stress and of rate of shear respectively. The ratios of deltas given in the fourth columns of these tables do not differ too greatly at any one value of shearing stress or of rate of shear. This fact shows that plots of log viscosity *versus* reciprocal temperature are reasonably linear. This experimental near linearity of log viscosity *versus* reciprocal temperature gives some support to the use of apparent activation energies to represent the variation of viscosity with temperature for these non-Newtonian systems. The apparent flow activation energies obtained from the slopes of such plots are also listed in Tables III and IV. The values of E_s^* calculated at different fixed shearing stresses are all about the same. These values are a little lower than the value that would be expected at extremely low shearing stresses, namely, around 12 kcal, but the difference is not large. The value of $E^*_{\dot{\gamma}}$ decreases with increase in the rate of shear at which it is calculated.

From all the above experimental data it can be generalized that for materials of the type concerned, the variation of viscosity with temperature at any fixed value of shearing stress is about the same as the variation at extremely low shearing stress, but that the variation at any finite fixed value of rate of shear is always lower than that at low shearing stress and decreases as one increases the rate of shear at which it is calculated. This conclusion goes further than the implications of the definition of viscosity, discussed in the preceding section, which merely showed that the variation at fixed shearing stress should always be greater than that at fixed rate of shear but said nothing about the absolute magnitude of either of them.

IV. EXAMINATION OF HYPERBOLIC SINE FLOW EQUATION

The experimental conclusions in the preceding section provide a criterion against which theoretical equations proposed to express non-Newtonian behavior of the type concerned can be checked. Specifically, if any theoretical equation proposed to express the non-Newtonian viscous behavior of high polymeric systems is such that variations of viscosity with temperature at fixed shearing stress and at fixed rate of shear can be derived from it, then these derived variations should

[6] See, e.g., T. G. Fox and P. J. Flory, J. Phys. & Colloid Chem. **55**, 221 (1951).

[7] Cragg, Faichney, and Olds, Can. J. Research **B26**, 551 (1948).

show the type of behavior found in the preceding section. To inaugurate a program of checking various existing theoretical equations against this criterion, the hyperbolic sine flow equation of Eyring[8] has been selected.

This equation is examined here because it is the basis for attempts to develop more adequate theoretical expressions. Theoretical expressions developed by Grunberg and Nissan,[9] by Bondi,[10] and by Spencer[11] all start with a simplified version of the hyperbolic sine equation and proceed by taking account of some process which is accepted as occurring in the type of non-Newtonian behavior with which the development is concerned, but which is not accounted for in the hyperbolic sine equation. The expressions of Grunberg and Nissan and of Bondi apply to ordinary liquids at high rates of shear and not specifically to high polymer systems. Spencer's expression seems to be the only one which has been developed specifically for such systems. The examination of these three last mentioned equations will be more directly pertinent than that of the hyperbolic sine equation, but since the latter is the antecedent of the more adequate expressions its examination is carried out here as the logical first step in the examination of the whole field.

Eyring's rate process theory develops rate of shear as

$$\dot{\gamma}=(2\lambda/\lambda_1)(kT/h)e^{-(\Delta H-T\Delta S)/RT}\sinh(s\lambda_2\lambda_3\lambda/2kT), \quad (11)$$

where all quantities on the right except s and T are considered as constants which need not all be defined here. The corresponding expression for viscosity is

$$\eta=(s\lambda_1/2\lambda)(h/kT)e^{(\Delta H-T\Delta S)/RT}\operatorname{csch}(s\lambda_2\lambda_3\lambda/2kT). \quad (12)$$

The usual Eyring equation for Newtonian viscosity is obtained by applying the approximation that for very small arguments of the csch, the csch is approximated by the reciprocal of its argument. The variation of viscosity with temperature in the resulting simplified equation is conveniently expressed by $d\ln\eta/d(1/T)$, which is a constant equal to $\Delta H/R$ if ΔH does not vary with temperature.

Differentiating Eq. (12) with respect to $1/T$ at constant s gives

$$E_s^*/R\equiv[\partial\ln\eta/\partial(1/T)]_s=\Delta H/R+T-(s\lambda_2\lambda_3\lambda/2k)\coth(s\lambda_2\lambda_3\lambda/2kT). \quad (13)$$

The quantity E^*_γ can be similarly evaluated by rewriting Eq. (11) as

$$s=(2kT/\lambda_2\lambda_3\lambda)\sinh^{-1}(\dot{\gamma}\lambda_1/2\lambda)(h/kT)e^{(\Delta H-T\Delta S)/RT}, \quad (14)$$

and thus getting

$$E^*_{\dot{\gamma}}/R\equiv[\partial\ln\eta/\partial(1/T)]_{\dot{\gamma}}=-T+(\Delta H/RT+1)\times(2kT^2/s\lambda_2\lambda_3\lambda)\tanh(s\lambda_2\lambda_3\lambda/2kT). \quad (15)$$

[8] See, e. g., Glasstone, Laidler, and Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York, 1941), p. 483.
[9] L. Grunberg and A. H. Nissan, Nature **156**, 241 (1945).
[10] A. Bondi, J. Appl. Phys. **16**, 539 (1945).
[11] R. S. Spencer, J. Polymer Sci. **5**, 591 (1950).

TABLE V. Temperature variation of viscosity according to hyperbolic sine flow law at $s=10^6$ dynes/cm².

T, °K	$[\partial\ln\eta/\partial(1/T)]_s$	$[\partial\ln\eta/\partial(1/T)]_{\dot{\gamma}}$
200	4596	1521
250	4639	1897
300	4676	2250
350	4708	2570
400	4736	2853
450	4759	3101

Both Eqs. (13) and (15) are consistent with the Newtonian approximation in that for either of them the right-hand side reduces to $\Delta H/R$ if one applies the approximation that at very small arguments of the tanh, the tanh is approximated by its argument. Equation (15) has the disadvantage that it gives $E^*_{\dot{\gamma}}$ as a function of s rather than $\dot{\gamma}$ as the independent variable. However, an expression for $E^*_{\dot{\gamma}}$ using $\dot{\gamma}$ as the independent variable would be extremely more complicated than Eq. (15), and Eq. (15) is perfectly useful if one keeps in mind that a value of $E^*_{\dot{\gamma}}$ obtained from it applies to whatever value of $\dot{\gamma}$ occurs for a given material at given values of s and of T.

It is evident from both Eqs. (13) and (15) that $\partial\ln\eta/\partial(1/T)$ decreases as s is increased, and in fact becomes negative if s is increased sufficiently. A previously stated requirement from the definition of viscosity now demands that both types of variations become negative (go through zero) at the same point. This point can be evaluated by making some numerical substitutions, for which we choose the following:

$$\lambda_2\lambda_3\lambda=V/N,$$
$$V=100\,000\ \text{cm}^3,$$
$$N=\text{Avogadro's number},$$
$$H=10\,000\ \text{cal},$$
$$T=300^\circ\text{K}.$$

The result is that both $[\partial\ln\eta/\partial(1/T)]_s$ and $[\partial\ln\eta/\partial(1/T)]_{\dot{\gamma}}$ go through zero at $s=8.8\times10^6$ dynes/cm². The change of sign in this case probably signifies that the flow equation no longer pretends to represent the viscous behavior in the region where this occurs. The equation probably represents only that part of the viscous behavior where deviations from Newtonian behavior are increasing, in which case it should not be given serious consideration at shearing stresses high enough so that the mechanisms responsible for the deviations are approaching saturation. At and somewhat before the point where the variations become zero, one is probably in the region of decreasing deviations, and hence should not attach much significance to the equation. It is known, however, that a shearing stress around 10^7 dynes/cm² is too low to be representative of this region for a material such as the hypothetical one considered here. This discrepancy is one illustration of the inadequacy of the hyperbolic sine equation itself.

The constancy of $\partial \ln\eta/\partial(1/T)$ with temperature, i.e., the linearity of $\ln\eta$ *versus* $1/T$, can be examined from Eqs. (13) and (15). Table V shows the values of $\partial \ln\eta/\partial(1/T)$ calculated at several different temperatures from Eqs. (13) and (15). These values were obtained using the numerical values stated previously, except for temperature, and taking a shearing stress of 10^6 dynes/cm^2, which is in the middle of the non-Newtonian region predicted by Eq. (12). It is seen immediately that this variation at fixed s is practically constant with temperature. The variations at fixed $\dot{\gamma}$ are not so easy to interpret since the fixed values at which they are calculated are different for the different temperatures. However, as the temperature is increased at a fixed shearing stress, the rate of shear increases, and as the rate of shear increases the variation calculated by fixing the rate of shear ($E^*_{\dot{\gamma}}$) decreases. Therefore, if the variations at fixed $\dot{\gamma}$ were calculated at the same fixed rate of shear for each temperature, they would change less with temperature than shown in Table V, and might easily be nearly constant in agreement with Table IV.

The values of $\partial \ln\eta/\partial(1/T)$ calculated at several different fixed values of shearing stress and rate of shear are shown in Fig. 4. The previously stated numerical values apply here. The values at fixed shearing stress (s^*) apply to the values of shearing stress given on the abscissa. The values at fixed rate of shear ($\dot{\gamma}^*$) apply at values of rate of shear corresponding to the values of shearing stress shown on the abscissa at the temperature used. As discussed previously, the values at the highest shearing stresses shown, say somewhat above 10^6, can probably be discounted as not pretending to represent the viscous behavior. However, in the region from somewhat above 10^4 to somewhat above 10^6 it seems justified to compare the behavior shown with the experimental behavior observed. In this region the theoretical behavior does in fact agree with the experimental behavior in that the variation at fixed shearing stress is but little different from the value at low shearing stresses, whereas the variation at fixed rate of shear decreases significantly as one increases the fixed rate of shear at which it is calculated.

V. CONCLUSION

The qualitative experimental knowledge of non-Newtonian viscous behavior of the type exhibited by high polymeric materials requires that the variation of viscosity with temperature at fixed shearing stress be greater than that at fixed rate of shear. However, this consideration says nothing about the absolute magnitude of either of these two types of variations.

Quantitative experimental data on several different systems of linear amorphous high polymers, both solvent-free and in solution, indicate that the variation at fixed shearing stress is always about the same as the variation at very low shearing stresses, whereas the variation at fixed rate of shear is always smaller than that at very low stresses, and decreases as one increases the rate of shear at which it is calculated.

The hyperbolic sine flow equation of Eyring, on which several attempts to develop theoretical flow equations are based, agrees with the nature of the above findings but gives too low values for the shearing stresses at which critical characteristics of the behavior are to be found. Both this flow equation and experimental data also suggest that apparent flow activation energies calculated either at fixed shearing stress or at fixed rate of shear are approximately constant with temperature.

These findings furnish an experimental criterion against which theoretical approaches to non-Newtonian flow for such materials can be checked. The relationship found between the variations at fixed shearing stress and at fixed rate of shear probably imply that as the temperature is varied the essential nature of the flow remains the same as long as the shearing stress is held fixed, but changes if the rate of shear is held fixed and the shearing stress allowed to vary. These considerations should be helpful in eventually arriving at a sound fundamental understanding of the nature of such non-Newtonian flow.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Effects of Rate of Shear on Inherent and Intrinsic Viscosities of Polystyrene Fractions*

L. J. SHARMAN,† R. H. SONES,‡ AND L. H. CRAGG
Department of Chemistry, Hamilton College, McMaster University, Hamilton, Ontario, Canada
(Received March 23, 1953)

Fractions of polystyrene, varying in molecular weight from 0.3×10^6 to 10×10^6 approximately, were prepared by careful tertiary fractionation. Inherent viscosities were determined at various rates of shear in the approximate range 100–20 000 sec^{-1}, viscosity measurements being made at several different temperatures in benzene and toluene (good solvents for polystyrene), in butanone (an intermediate solvent), and in cyclohexane (a poor solvent). From these values intrinsic viscosities at various rates of shear were determined. These viscosity functions are sensitive to shear rate at high molecular weights; at a given temperature the shear effects are greater the higher the molecular weight and the better the solvent, and with increasing temperature they decrease in a good solvent and increase in a poor one This behavior is shown to be consistent with current ideas regarding the size and shape of polymer molecules in solution as affected by weight, temperature, and solvent power. It is concluded, among other things, that with flexible polymers (or polymer fractions) of very high molecular weight, measurements of intrinsic viscosity should be made in a poor solvent at a low temperature.

INTRODUCTION

YEARS ago it was recognized that a solution of a high polymer in an organic solvent may be a non-Newtonian liquid.[1] Nevertheless, in the great majority of recent investigations involving the evaluation of intrinsic viscosity, the rate of shear has not been treated as a significant variable even though in capillary viscometers as ordinarily used the rate of shear may be altered considerably by a change in solvent, in molecular weight, or in temperature. It is true that the effects become negligible when the molecular weight of the fraction and the concentration of the solution are low enough, but the most interesting polymers are not always those of low molecular weight, nor are the most interesting solutions always the very dilute ones. Indeed with the recent developments in the theoretical treatment of intrinsic viscosity[2] and with the availability of better methods for determining very high molecular weights, interest in the very high polymers is rapidly increasing. When molecular weights get high enough, the rate-of-shear effects can no longer be ignored with impunity, even when measurements are made in the most dilute solutions practicable.

A number of investigators have recently been concerning themselves with these rate-of-shear effects both experimentally[3] and theoretically,[4] but there is still considerable disagreement as to how to take account of them in determining such viscosity functions as the intrinsic viscosity $[\eta]$, the problem being, of course, that the magnitude of the effect varies with the rate of shear. Most investigators have favored an extrapolation either to infinite rate of shear[5] or, more commonly, to zero rate of shear.[6] Others advocate making all measurements at some arbitrary but finite rate of shear.[7] It is clear that much more work needs to be done before this issue can be decided. Accordingly, we have made careful measurements of rate-of-shear effects on inherent viscosity and intrinsic viscosity of polystyrene fractions and of the influence on these effects of changing the molecular weight of the polymer, the temperature of the solutions, and the solvent used.

EXPERIMENTAL METHODS

A. Materials

The *solvents* were middle fractions of constant refractive index obtained by slow careful distillation through a 3-ft column packed with glass helices. Other details were as follows:

benzene: (Steel Company of Canada) ASTM Industrial Grade, dried over sodium, $n_D^{25}=1.4975$;
toluene: (Steel Company of Canada) Nitration Grade, dried over sodium, $n_D^{25}=1.4922$;
butanone (MEK): (Shell Oil Company), Technical Grade, dried over anhydrous calcium sulfate ("Drierite"), $n_D^{25}=1.3755$;
cyclohexane: (Fisher Scientific Company) C. P. Grade; dried with, distilled from, and kept over sodium; $n_D^{25}=1.4233$.

* Presented at the Annual Meeting of the Society of Rheology, Philadelphia, Pennsylvania, October 30, 31, 1952.
† Present address: Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan.
‡ Present address: Canadian Industries Limited, Nylon Division Kingston, Ontario, Canada.
[1] H. Staudinger, *Die hochmolekularen organischen Verbindungen* (Springer, Berlin, 1932), p. 189.
[2] P. Debye and A. M. Bueche, J. Chem. Phys. **16**, 573 (1948); J. G. Kirkwood and J. J. Riseman, J. Chem. Phys. **16**, 565 (1948); **17**, 442 (1949); P. J. Flory and T. G. Fox, Jr., J. Am. Chem. Soc. **73**, 1904 (1951).
[3] See, as examples, Conrad, Tripp, and Mares, J. Phys. & Colloid Chem. **55**, 1474 (1951). R. M. Fouss and W. N. Maclay, J. Polymer Sci. **6**, 305 (1951); Fox, Fox, and Flory, J. Am. Chem. Soc. **73**, 1901 (1951); Akkermans, Pals, and Hermans, Rec. trav. chim. **71**, 56 (1952).
[4] See, as examples, R. Simha, Proc. Intern. Congr. Rheol. Holland, 1948, II-68; G. de Wind and J. J. Hermans, Rec. trav. chim. **70**, 521 (1951); A. Peterlin, J. Polymer Sci. **8**, 621 (1952); and papers cited therein.
[5] C. N. Davies, Trans. Faraday Soc. **40**, 274 (1944).
[6] For example, Fox, Fox, and Flory, reference 3.
[7] Tripp, Conrad, and Mares, J. Phys. Chem. **56**, 693 (1952).

TABLE I. Intrinsic viscosities of the fractions at zero rate of shear and molecular weights calculated from various equations of the form $[\eta]=KM^a$.

Fraction	$[\eta]_{D=0}$ in toluene at 30°C	$M(\times10^{-6})$ F & F[a]	B,F & K[b]	O,C & M[c]	G,H & M[d]	F & B[e]
T_5	11.8	9.2	9.2	10.3	18.4	3.3
C_1A_1	9.2	6.5	6.5	7.3	12.3	2.4
F_1A_1	8.1	5.5	5.5	6.1	10.1	2.1
T_{15}	7.8	5.2	5.2	5.8	9.4	
T_3	4.40	2.4	2.3	2.5	3.7	
F_5A_1	4.20	2.2	2.2	2.3	3.5	0.86
T_4	3.11	1.5	1.4	1.5	2.1	
F_7B_1	2.75	1.2	1.2	1.3	1.8	0.49
F_7B_3	1.80	0.70	0.65	0.68	0.89	0.28
F_3C_2	0.97	0.30	0.27	0.28	0.33	0.12

[a] Equation of Fox and Flory (reference 10, rearranged): $[\eta]=0.98\times10^{-4}M^{0.78}$.
[b] Equation of Bawn, Freemen and Kamaliddin (reference 11): $[\eta]=1.34\times10^{-4}M^{0.71}$.
[c] Equation of Outer, Carr, and Zimm (reference 12): $[\eta]=1.7\times10^{-4}M^{0.69}$.
[d] Equation of Goldberg, Hohenstein, and Mark (reference 13): $[\eta]=3.70\times10^{-4}M^{0.62}$.
[e] Equation of Frank and Breitenbach (reference 14): $[\eta]=1.3\times10^{-4}M^{0.76}$.

The *polystyrene* fractions were from two sources. Those labeled F_3C_2, F_7B_1, F_5A_1, $F_1A_1A_1$, and C_1A_1 derived from polystyrene prepared by emulsion polymerization at 55°C to 96 percent conversion. They were products of a very careful three-stage fractionation and are believed to be unusually "sharp" and free of branched species.[8] Those labeled T_3, T_4, T_5, and T_{15} were likewise very "close" tertiary fractions;[9] the original sample of polystyrene from which they were separated was prepared by allowing styrene, free of catalysts or inhibitors, to polymerize slowly in bulk at room temperature over a period of two years.

The molecular weights of these fractions were calculated from intrinsic viscosities (at zero rate of shear) in toluene at 30°C, using the equation of Ewart and Tingey as modified by Fox and Flory,[10] and are recorded in Table I. For purposes of comparison, values were also calculated—see Table I—from the appropriate intrinsic viscosities using certain other equations of the many available, namely, those of Bawn, Freeman, and Kamaliddin,[11] Outer, Carr, and Zimm,[12] Goldberg, Hohenstein, and Mark,[13] and Frank and Breitenbach.[14]

Fortunately we are concerned in this paper more with relative than with absolute values, particularly in the high molecular weight range. (The four highest values especially are to be regarded with suspicion, partly because in this range extrapolation to zero rate of shear is dubious and partly because the equations are here being applied beyond the range to which validity was claimed for them.)

[8] J. A. Manson and L. H. Cragg, Can. J. Chem. **30**, 482 (1952).
[9] R. H. Sones, M. Sc. thesis, McMaster University, Hamilton, Ontario, Canada, 1952.
[10] T. G. Fox, Jr., and P. F. Flory, J. Am. Chem. Soc. **73**, 1915 (1951).
[11] Bawn, Freeman, and Kamaliddin, Trans. Faraday Soc. **46**, 1107 (1950).
[12] Outer, Carr, and Zimm, J. Chem. Phys. **18**, 830 (1950).
[13] Goldberg, Hohenstein, and Mark, J. Polymer Sci. **2**, 503 (1947).
[14] H. P. Frank and J. W. Breitenbach, Monat. Chem. **81**, 570 (1950).

B. Apparatus

The *viscometer* was of the Ostwald type but was specially designed for the investigation. The maximum rate of shear in a capillary viscometer (at the wall) is given by the equation

$$D=rghd/2l\eta, \tag{1}$$

where r is the radius of the capillary and l its length, d is the density of the liquid and η its (apparent) viscosity, g is the acceleration due to gravity, and h is the effective hydrostatic head. During a measurement the head decreased; the *mean rate of shear* is obtained by using in Eq. (1) the mean head, i.e., the head at $t/2$ where t is the total time of flow. From Eq. (1) it may be seen that for a given liquid the mean rate of shear may be varied by varying one or more of r, h, and l. It is, however, convenient to use the same viscometer for all measurements, i.e., to keep r and l constant and vary h by applying an additional (external) pressure to one or other of the arms of the viscometer.

The modified Poiseuille equation may be written

$$\eta=\frac{\pi r^4hgdt}{8V(l+nr)}-\frac{mdV}{8\pi t(l+nr)}, \tag{2}$$

where V is the volume of the efflux bulb and t the time required for this volume of liquid to flow through the capillary, and n and m are constants. The term nr is the end correction ($l+nr$ being the effective length) and the term $mdV/(8\pi t(l+nr))$ is the kinetic-energy correction. These corrections will be minimized by making l large and r and v small. A large ratio, l/r, will also permit the easier attainment of very low rates of shear [see Eq. (1)].

Accordingly, the viscometer was provided with a long fine capillary and a small efflux bulb. In addition, a short section of the tube between the capillary and the bulb was of wider radius to minimize surface tension effects.

Equation (2) has the form

$$\eta=Adt-(Bd/t) \tag{3}$$

or

$$\nu=\eta/d=At-(B/t), \tag{4}$$

where ν is the "kinematic viscosity," and A and B are instrument constants for a particular viscometer. (It should be emphasized that frequently, because of imperfect construction, the values of A and B are not constant, but vary with the flow time t.) Values of A and B were determined by calibration with pure benzene and toluene at several different temperatures. Equation (4) may be written in the form

$$\nu/t=A-B(1/t^2). \tag{5}$$

Hence, from a plot of ν/t *vs* $1/l^2$, A is obtained as an intercept and B as a slope. Values of ν were calculated from values of η and d obtained by interpolation from the data of Timmermanns[15] and the values of t were experimental. From the plot obtained it was clear that for this viscometer B is constant over the whole range of our measurements. The actual values of B and A are given in Table II along with other constants of the viscometer. Rearranging Eq. (4) again

$$\frac{\nu}{A}=t-\frac{B}{A}\cdot\frac{1}{t}.$$

The kinetic-energy correction, $-(B/A)\cdot 1/t$, would be greatest for the shortest flow times. B is so small for our viscometer that in the shortest flow time measured in our experiments, 7 sec, the correction is only 0.005 sec. Hence, the kinetic-energy effect could be—and was—neglected in all the measurements reported in this paper. When the measurements with a polymer solution and with the pure solvent are all made in this viscometer $\nu_s/\nu_0=t_s/t_0$, where subscripts s and 0 refer to solution and solvent, respectively. Since, moreover, the densities of the dilute polystyrene solutions in these experiments differ only very slightly from those of the solvents used in preparing them, d_s/d_0 may be taken as unity, and $\nu_s/\nu_0=\eta_s/\eta_0$. Thus, relative viscosities can be determined directly as relative flow times.

The *manostat*, by means of which an external pressure could be set, maintained, and applied to the liquid in the viscometer, was similar to that described by Fuoss and Cathers[16] but modified in two ways: (1) by introducing between the main body of the manostat and the viscometer a five-gallon well-insulated bottle containing air, to minimize pressure variations particularly at low pressures, and (2) by introducing between this bottle and the viscometer a drying tube. By means of this manostat, external pressures could be maintained constant throughout the period of measurement within ± 0.1 cm of water. By applying the pressure to the appropriate arm of the viscometer (i.e., using it either to hasten or retard the flow of liquid) the amount by which the effective head was increased was varied from -17 to $+140$ cm.

Constant temperature baths were maintained at 15, 25, 45, 65, and 85°C, the temperature in each being controlled to ± 0.02°C (more closely at the lower temperatures). Each was fitted with a clamp incorporating a standard-taper socket fitting a pin on the clamp holding the viscometer. By this means the viscometer could be moved from one bath to another to reproducibly-vertical positions.

[15] J. Timmermans, *Physico-chemical Constants of Pure Organic Compounds* (Elsevier Publishing Company, Inc., New York, 1950).
[16] R. M. Fuoss and G. I. Cathers, J. Polymer Sci. **6**, 301 (1951).

C. Procedure

Preparation of Solutions

The most concentrated solution of each run was prepared by dissolving a weighed amount of the polystyrene in solvent, with gentle swirling and diluting to volume at 25.0°C. Other solutions were prepared by diluting a measured volume of this solution with a measured volume of solvent, both at 25.0°C.

Concentrations were calculated in the units g/100 ml solution, at 25.0°C from the known weight and volume, at other temperatures by multiplying by the density ratio d_t/d_{25}. (Densities of the solvent rather than of solution were used but this was shown to introduce negligible error.)

Viscosity Determinations

The viscometer, into which 10 ml of solution had been pipetted after filtration through fritted-glass, was placed in the 25°C bath and left for 15 minutes to insure thermal equilibrium. Flow times were then measured at six different pressures. In the same way measurements were then made at the other temperatures, the viscometer being moved successively to the baths at 15°, 45°, 65° and 85°C. (Even after this long series of measurements, evaporation was so slight as to cause no change in the flow time of the solution at 25°.) Flow times were measured, usually in triplicate, with accurate calibrated stopwatches reading to 0.05 sec. The precision of measurement varied from ± 0.02 sec at 8 sec to ± 2.0 sec at 2000 sec; this corresponded to a precision of ± 5 percent in the intrinsic viscosity at the slowest and fastest flows, falling rapidly to ± 1 percent when t_s was in the range 50–400 sec.

TABLE II. Constants of the viscometer.

l	16.9 cm	V	0.83 ml
r	0.0177 cm	A	5.819×10^{-3}
h(at $t/2$)	22.9 cm	B	2.00×10^{-4}

Calculation of Rate of Shear

The mean rate of shear at the wall was calculated using Eq. (1) in the form

$$D=rghd/2l\eta_r\eta_0.$$

Values of η_r were experimentally determined, as t_r. Values of η_0 and d were obtained by interpolation from the data of Timmermans (d being taken as equal to d_0). Values of h, the effective hydrostatic head, were obtained by adding the applied pressure to (or subtracting it from) the mean hydrostatic head in the viscometer (i.e., the head at $t_0/2$), both being first converted to the units dynes/cm^2.

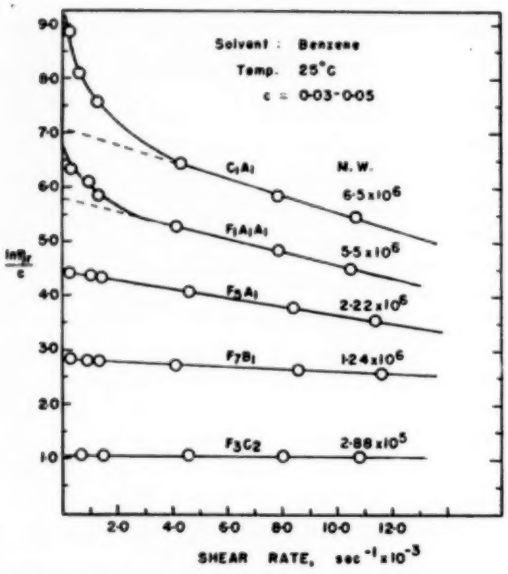


FIG. 1. Typical inherent viscosity *vs* shear-rate curves for five polystyrene fractions of different molecular weight (at approximately equal concentrations).

RESULTS

In a given solvent several (4 or 5) solutions of each fraction were prepared, the concentrations covering about a tenfold range, down to about 0.03 g/100 ml for the lower fractions, and to about 0.01 g/100 ml for the higher fractions in a good solvent. With each of these solutions, flow times were measured at 6 or 8 different shear stresses (i.e., settings of the manostat) the whole set of measurements being repeated at several different temperatures. The data are, therefore, far too numerous to record here.

The data so obtained were used as follows: Values of $\ln\eta_r/c$ were calculated as $\ln t_r/c$ for each rate of shear These values were then plotted against the shear rate D and a smooth curve drawn through the points. (Typical examples are shown in Figs. 1, 2, and 3.) From this curve were obtained, by extrapolation, $(\ln\eta_r/c)_{D=0}$ the inherent viscosity at zero rate of shear, and by interpolation $(\ln\eta_r/c)_{D=D'}$ the inherent viscosity at any desired finite rate of shear D' (within the range of measurements, of course).

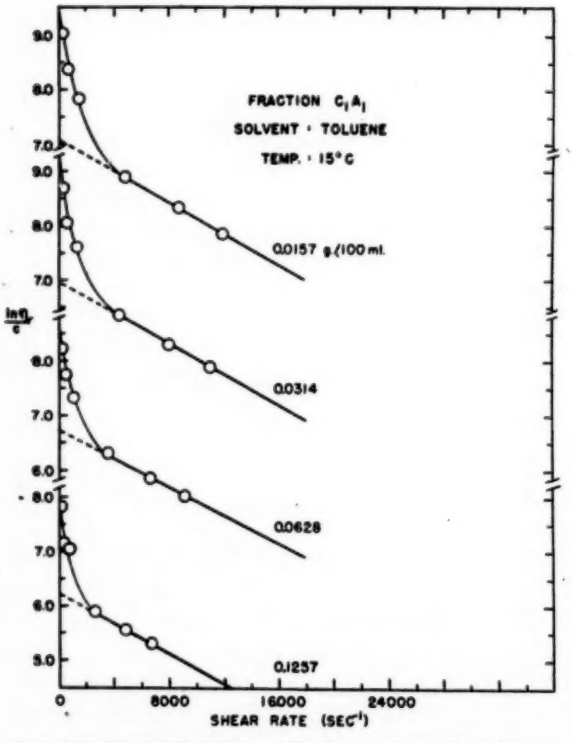


FIG. 2. Inherent viscosity *vs* shear-rate curves at 15°C, for one polystyrene fraction at four different concentrations.

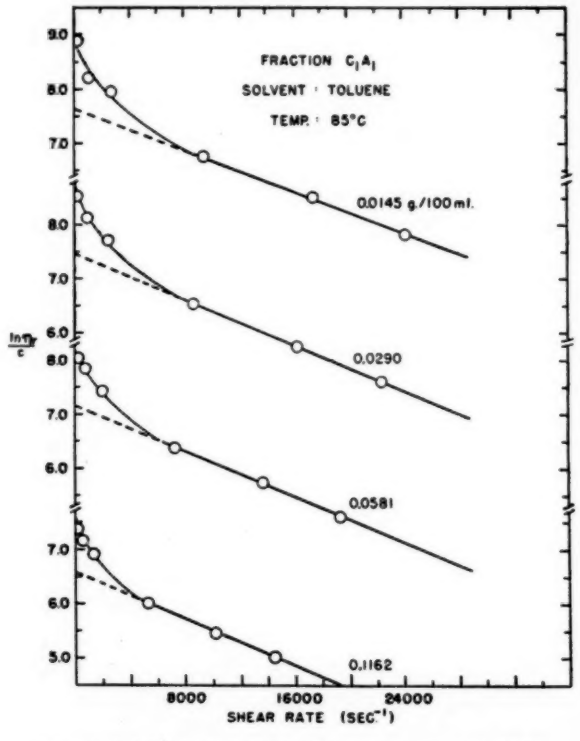


FIG. 3. Inherent viscosity *vs* shear-rate curves at 85°C. (The solutions are the same as those of Fig. 2).

The values of inherent viscosity so obtained for each concentration (for a given fraction in a given solvent at a given temperature) were then plotted against concentration and the corresponding intrinsic viscosity, $[\eta]_{D=0}$ or $[\eta]_{D=D'}$ as the case might be, were obtained by extrapolation to $c=0$ in the usual way (see, for example, Fig. 4).

Since measurements were made with several fractions, in four different solvents (varying from very good to very poor), at several different temperatures, and over a range of shear rates, such calculations as these provided the data for determining not only the effect on inherent viscosity and on intrinsic viscosity of changes in each of these variables but also the influence of molecular

weight, temperature, and solvent power on the shear effects where these were evident.

It is worth emphasizing that there were enough measurements to make it possible to determine these effects with several independent sets of data, and hence conclusions could be drawn with much greater confidence.

From our data, then, it is evident that at constant temperature and with a polymer fraction of high enough molecular weight:

(1) The *inherent viscosity* at any given (low) concentration decreases with increasing rate of shear; and the rate of this decrease is greater (a) the higher the molecular weight (Fig. 1), (b) the better the solvent (Fig. 5), and (c) the lower the temperature in a good solvent (Fig. 6) and the higher the temperature in a poor solvent (compare Figs. 7 and 8).

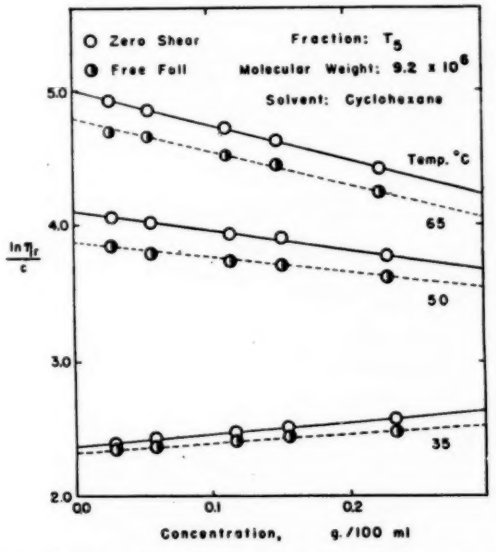


FIG. 4. Typical inherent viscosity *vs* concentration curves. ("Free fall" measurements were made in the usual way, i.e., with liquid falling under its own weight.)

(2) The *intrinsic viscosity* likewise decreases with increasing rate of shear; and the rate of this decrease is greater (a) the higher the molecular weight (Fig. 9), (b) the better the solvent, and (c) the lower the temperature in a good solvent (Fig. 10) and the higher the temperature in a poor solvent (Fig. 11).

Actually, there appear to be *two* effects of shear on the inherent viscosity, one superimposed on the other. In Fig. 1, which is quite typical of the curves for a series of fractions in a good solvent at one temperature, the curve for the lowest fraction, F_3C_2 is horizontal, indicating no effect of shear rate. With the next two fractions, F_7B_1 and F_5A_1, there is an effect—a rectilinear decrease in inherent viscosity with increasing shear rate. With the two top fractions, $F_1A_1A_1$ and C_1A_1, there is a second effect superimposed on this first one, namely, a curvilinear decrease. Putting it another

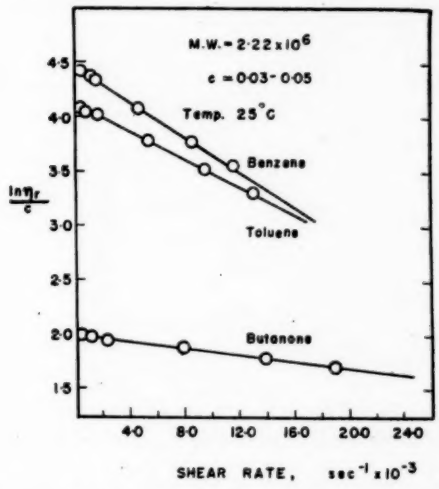


FIG. 5. Effect of solvent on inherent viscosity *vs* shear-rate curves.

way, the points corresponding to higher shear rates fall on a straight line but at lower and lower shear rates the points are further and further above the extension of this straight line. A great many curves of this type, i.e., with both a straight and curved portion, were obtained (for further examples see Figs. 2 and 3), and hence the evidence for the two different effects is strong. As the molecular weight of the fraction increases, however, the curved portion extends to higher and higher shear rates, until eventually the line is curved over the whole experimental range. (See Fig. 12.)

The magnitude of these two effects can be estimated, at least roughly, whenever a straight-line portion of the $\ln\eta_r/c$ *vs.* D curve can be identified. For effect I a suitable measure is the slope, $d(\ln\eta_r/c)/dD$, of the straight line. For convenience, we shall use for this

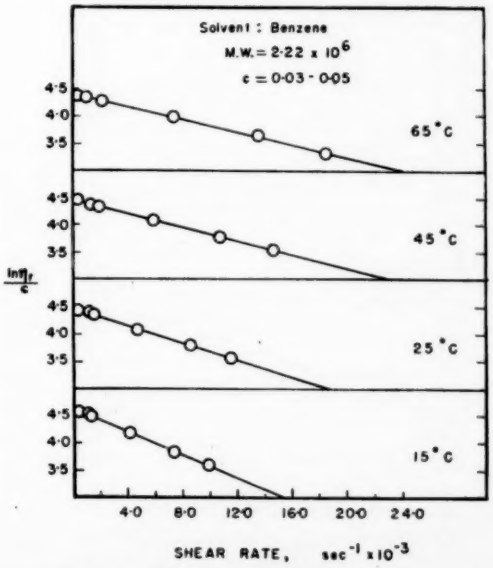


FIG. 6. Effect of temperature on inherent viscosity *vs* shear-rate curves in a good solvent.

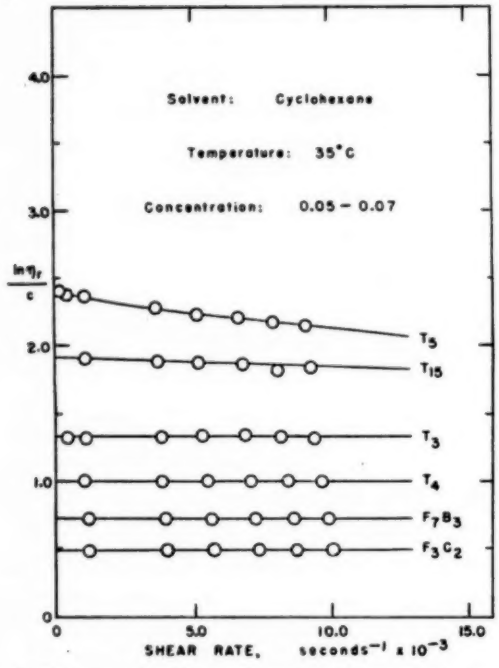


FIG. 7. Inherent viscosity *vs* shear-rate curves for six fractions in cyclohexane at 35°C.

slope the symbol J. For effect II the choice of a measure is much more difficult. The most direct would seem to be the difference, Δ_{iv} on Fig. 13, between the value of $\ln\eta_r/c$ at $D=0$ obtained by extrapolating the curved portion and that obtained by extrapolating the linear portion. This suffers from the obvious defect that the question of what happens at very low rates of shear is still unanswered. In other words, one cannot yet extrapolate this curved portion with confidence. The measure is, therefore, both empirical and rough; nevertheless, as we shall see, it is still usable.

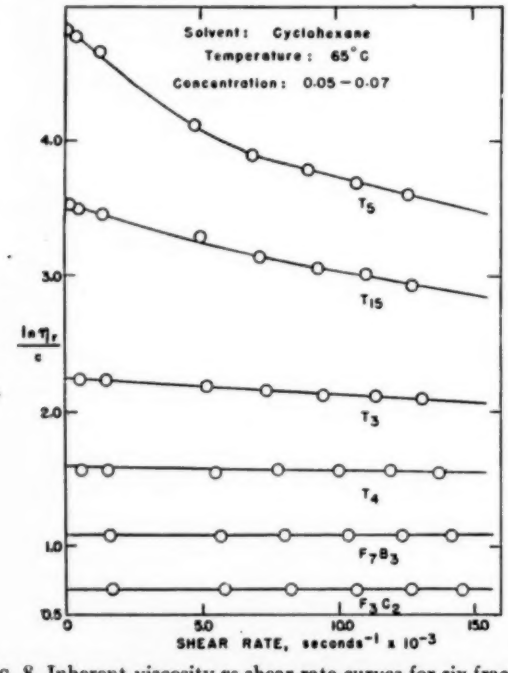


FIG. 8. Inherent viscosity *vs* shear-rate curves for six fractions in cyclohexane at 65°C.

With increasing molecular weight, effect I appears first, then effect II appears, and eventually becomes large enough to obscure the first effect entirely. It is also possible, therefore, to obtain a threshold value for each of the effects, e.g., the minimum value of molecular weight at which they are observable. Our experiments were not designed to determine these threshold values, but the data obtained often do make it possible to estimate them approximately in the four solvents and at several temperatures.

All of these semiquantitative conclusions may be briefly summarized.

I. The Rectilinear Decrease with Increasing Shear Rate

(*a*) *Threshold Value*

In benzene, toluene, and butanone the effect of shear rate is undetectable at $M=0.3\times10^6$ and appreciable at 1.2×10^6, the threshold value probably being in the neighborhood of 0.7×10^6 in all three solvents. In these three solvents this threshold value is affected little, if at all, by change of temperature in the range 15–85°C. In cyclohexane, however, the threshold value is probably below 1.0×10^6 at 65°C, but above 2.0×10^6 (almost 2.4×10^6) at 35°C (see Figs. 7 and 8).

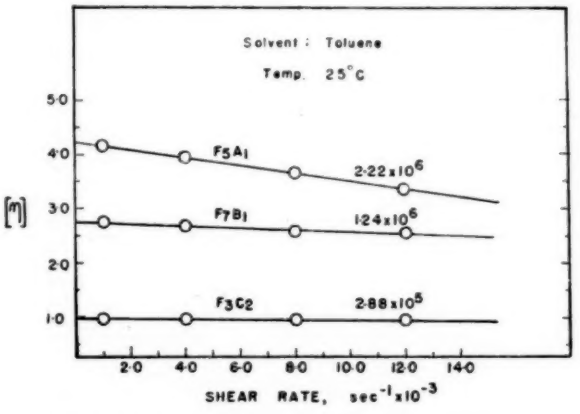


FIG. 9. Intrinsic viscosity *vs* shear-rate curves at three different molecular weights.

Expressed in terms of intrinsic viscosity the threshold value is, on the other hand, almost independent of temperature (see Figs. 7 and 8). Although the evidence on this point is somewhat scanty, the threshold value seems to be somewhat higher in the good solvents (approximate values being $2._0$ in toluene and in benzene, $1._0$ in butanone, and $1._3$ in cyclohexane).

(*b*) *The Slope J*, i.e., $d(\ln\eta_r/c)/dD$

The magnitude of the slope J is influenced by several factors.

(1) J increases with increasing molecular weight of the polymer.

(2) J increases with increasing "goodness" of the solvent (see Fig. 5). A solvent in which the inherent viscosity of a fraction (at a given shear rate, concentration, and temperature) is high is one in which this slope is high.

(3) J decreases with increasing temperature in a good solvent, such as benzene or toluene (compare Fig. 2 with Fig. 3), but it increases with increasing temperature in a very poor solvent like cyclohexane (compare Fig. 7 with Fig. 8).

(4) J is almost independent of concentration, but fairly consistently there appears to be a slight increase with increasing concentrations.

A small part of the evidence for these conclusions is assembled in Table III.

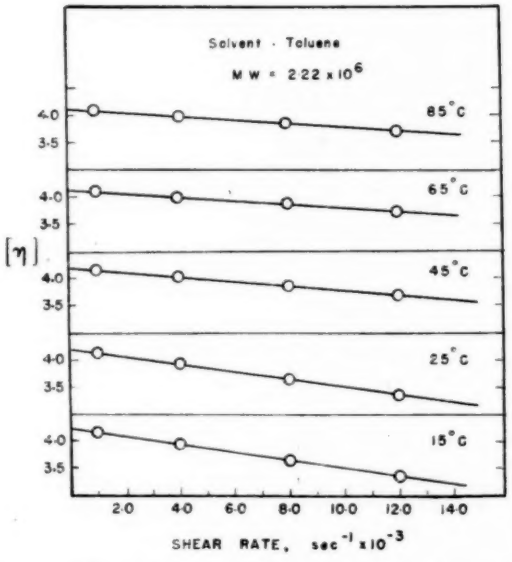


FIG. 10. Effect of temperature on the intrinsic viscosity *vs* shear-rate curve in a good solvent.

II. The Curvilinear Decrease with Increasing Shear Rate

(*a*) *Threshold Value*

No curvature was observed, in any solvent at any temperature, at or below $M=2.2\times10^6$. At $M=5\times10^6$, however, it was already considerable in the good solvents and was detectable in the poor solvent cyclohexane at the highest temperature 65°C though not at the two lower temperatures.

In terms of molecular weight the threshold value seems to be higher in a very poor solvent than in a good solvent. On the other hand, when expressed in terms of inherent viscosity (at some fixed shear rate) the threshold value is lower in the very poor solvent. (Compare Fig. 7 with Fig. 1).

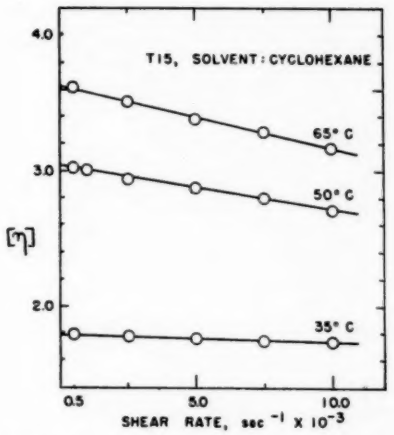


FIG. 11. Effect of temperature on the intrinsic viscosity *vs* shear-rate curve in a poor solvent.

(*b*) Δ_{iv}

As already stated, this measure of the second shear effect is very approximate. Nevertheless, certain trends are consistently manifested:

(1) Δ_{iv} increases with increasing molecular weight (Figs. 1 and 8).

(2) Δ_{iv} increases with increasing "goodness" of the solvent (Fig. 14).

(3) Δ_{iv} decreases with increasing temperature in a good solvent such as benzene or toluene (compare Figs. 2 and 3), but it increases with increasing temperature in a poor solvent like cyclohexane (compare Figs. 7 and 8).

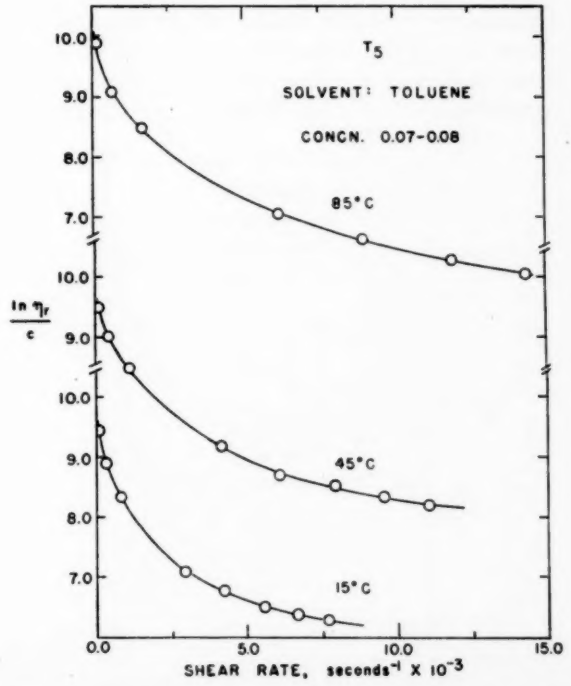


FIG. 12. Inherent viscosity *vs* shear-rate curves for a high-molecular-weight fraction in a good solvent at three temperatures.

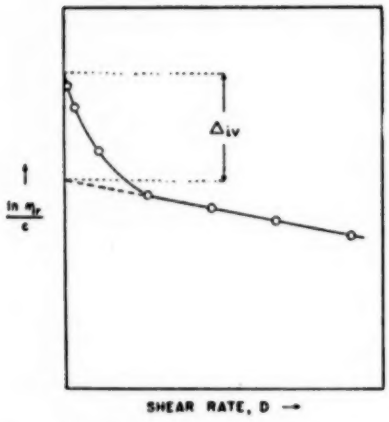


FIG. 13. Graphical definitions of the function Δ_{iv}.

(4) Δ_{iv} appears to decrease slightly with increasing concentration in a good solvent (Fig. 3) and to increase very slightly with increasing concentration in a very poor solvent (cyclohexane at 35°C).

DISCUSSION

It should be possible to explain all of these effects in terms of molecules and their behavior. The progressive decrease in apparent viscosity (and in the viscosity functions) with increasing shear, observed with many different polymer-solvent systems, has been attributed variously to increasing deformation of flexible polymer molecules, to increasing orientation of these molecules in the direction of flow, to progressive breakdown under shear of polymer-molecule associations (whether arising from intermolecular attractions or from entanglement) and to combinations thereof. One would expect an effect resulting from deformation and orientation of isolated polymer molecules to show up earlier (i.e., at lower molecular weights) than one resulting from interactions

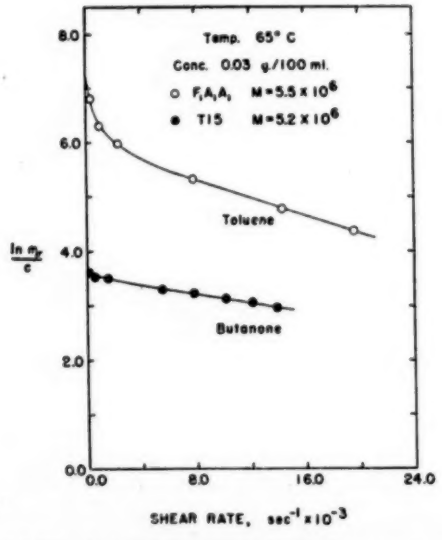


FIG. 14. The decrease in rate-of-shear effects in changing from a good to a poor solvent.

of polymer molecules. It is tempting, therefore, to attribute the rectilinear decrease in inherent viscosity to molecular deformation and orientation and the curvilinear decrease to molecular interactions. Indeed, if we do this we can find plausible explanations for most of the phenomena observed. But, if the second effect is due to molecular interaction, it should ultimately disappear or at least markedly decrease as the concentration is lowered. In spite of special efforts we were unable to observe such a decrease even at a molecular weight at which Δ_{iv} was small. This failure may, of course, be due to the difficulty of making meaningful measurements when both the concentrations of the solutions and the rate of shear are very low. (The length of a vertical line in Fig. 15 indicates roughly the precision of the value

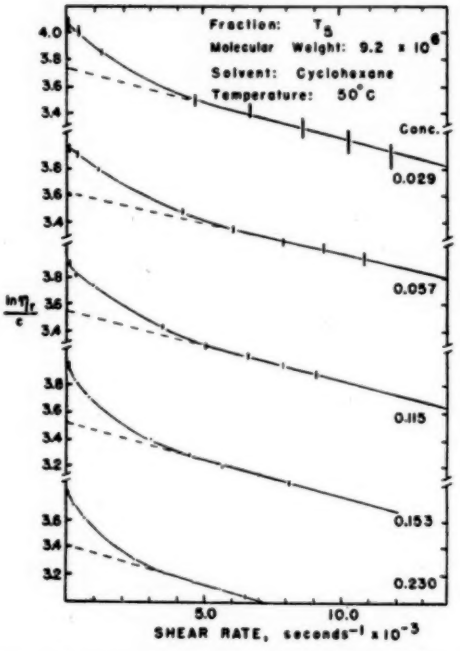


FIG. 15. Inherent viscosity *vs* shear-rate curves. Relative precision of measurement as affected by rate of shear and concentration of solution.

of $\ln\eta_r/c$ plotted.) Certainly, precise measurements at very low shear rates and very low concentrations are badly needed.

Nevertheless, there is good reason to believe that a satisfactory explanation for these shear phenomena can be found, because many of them can already be fitted into a consistent broad picture. It is generally accepted that a very large polymer molecule in solution assumes, as a result of random kinking, the form of a loose "ball"§ more or less expanded, depending on the nature of the solvent, on the temperature, and on the inherent flexibility of the polymer chain. Whether a given shear effect is due to molecular deformation and orientation or to interaction

§ We prefer this to the more usual descriptive term "coil," which suggests an ordered configuration. In all probability the molecule is not neatly coiled but, rather "all balled up."

of the polymer molecules, it should be greater when this "ball" is larger and looser. Whatever enlarges and expands the molecular "ball" should increase the shear effect. And this is just what is observed. Increasing the weight of the molecule, putting it in a better solvent, increasing the temperature of the solution if the solvent is poor and decreasing it if the solvent is good, all these make the effective volume of the molecule (the size of the molecular "ball" to which we have been referring) greater—and all these make both J and Δ_{iv} greater! It is evident, then, that the rate-of-shear effects are closely related to the effective volume of the polymer molecules in solution.

Regardless of explanations, there are three practical conclusions that can be drawn from these results. First, the effect of shear rate on the intrinsic viscosity of polystyrene is by no means negligible at molecular weights where hitherto it has customarily been neglected. Second, on such plots as $\ln\eta_r/c$ *vs* D, a curve may only begin to deviate from a straight line at a fairly low rate of shear, and hence straight-line extrapolation is safe only when measurements are made at as low rates as possible (and possibly not even then!). Third, the difficulties of extrapolation to zero rate of shear become greater the higher the molecular weight and the resulting values of $[\eta]_{D=0}$ become correspondingly less reliable.

TABLE III. Variations of the slope J with temperature, solvent, and concentration. (M.W. $= 2.2\times10^6$ in toluene and butanone, M.W. $= 2.4\times10^6$ in cyclohexane)

Toluene			Butanone			Cyclohexane		
Conc g/100 ml	$J(\times10^4)$ 15°C	65°C	Conc g/100 ml	$J(\times10^4)$ 15°C	65°C	Conc g/100 ml	$J(\times10^4)$ 35°C	65°C
0.039	0.64	0.38	0.018	0.17	0.15	0.035	0.01	0.10
0.079	0.72	0.39	0.036	0.16	0.16	0.071	0.01	0.12
0.157	0.76	0.40	0.071	0.15	0.14	0.143	0.02	0.13
0.315	0.78	0.42	0.143	0.15	0.15	0.189	0.04	0.14

The shear effects are, however, very greatly reduced in a very poor solvent at a low temperature (close to the precipitation temperature). This would suggest that with polymers of high molecular weight the intrinsic viscosity should be measured in a very poor solvent rather than in a very good one, as is now the custom. (We plan to elaborate on this point in a later publication.)

ACKNOWLEDGMENTS

We wish to express our gratitude for the financial assistance received from the Research Council of Ontario in the form of Scholarships and from the National Research Council (of Canada), through its Associate Committee on Synthetic Rubber Research, in the form of grants-in-aid.

Anomalous Viscous Flow at Very Low Rate of Shear and Small Shearing Stress

J. T. Bergen and W. Patterson, Jr.*
Research and Development Center, Armstrong Cork Company, Lancaster, Pennsylvania
(Received March 23, 1953)

Fluid suspensions of a solid phase in a liquid, for example, paints and printing inks, have been presumed to exhibit anomaly of flow at very low rate of shear. The consistency characteristics of such materials require an experimental shear rate and shear stress much lower than has heretofore been reported.

This paper describes a translated cylinder, or translational, viscometer of a modified Pochettino type, which possesses a shear stress sensitivity of 5×10^{-3} dyne/cm², and a shear rate sensitivity of 1×10^{-4} sec^{-1}. A rubber Latex of high solids content exhibited no anomaly of flow within this range of sensitivity. A rubber dissolved in organic solvent was observed to approximate a Bingham material of 0.05-dyne/cm² yield value, while a gum tragacanth suspension in water exhibited a yield value of 0.4 dyne/cm² with marked post-yield anomalous flow. Household paints exhibited yield value ranging from zero to 6 dyne/cm². These yield values were related to leveling-out of brushmarks, and to sagging or "curtain," except for examples exhibiting marked thixotropic behavior. Industrial organosol coatings likewise possessed yield values ranging up to 5 dyne/cm² which were related to leveling characteristics. These yield values agree with those postulated by Waring from consideration of the forces of leveling of brushmarks. They are considerably smaller than those reported in the literature from extrapolation of flow curves from much higher rate of shear.

INTRODUCTION

THE rheology of materials of technological interest which are known or presumed to exhibit anomalies of flow has been the subject of considerable investigation. Bingham,[1] for example, considers clay suspensions, and Green[2] discusses paints and printing inks. No generalized theory which will account for or predict anomalies of flow has yet appeared, particularly with respect to the case of solids suspended in a fluid. It is thus necessary that the flow characteristics of systems exhibiting anomalous or non-Newtonian flow properties be observed experimentally throughout the whole range of shearing rates or shear stresses which may be of technological significance. Houwink[3] has stated the problem very well:

"With regard to flow, the really important datum should be the whole course of the (shear rate-shear stress) curve so far as it has been deduced by exact methods from the most appropriate experimental investigations."

With regard to the low end of the shear rate-shear stress curve, experimental investigations have been carried out at rates of shear of the order of 0.001 sec^{-1}, but only for shear stress upwards of thousands of dynes/cm², i.e., only for very viscous materials.[4,5] On the other hand, for shear stresses of a few dynes/cm² the associated shear rates have been upwards of 0.1 sec^{-1},[6] such materials[3] are of quite low viscosity. Fluids of technological importance often fall between these two extremes; an investigation of the flow curve for such materials implies that the lower limits of accurate measurement of both shear stress and shear rate be kept as small as possible.

An example may be drawn from the analysis by Waring[7] of the magnitude of forces involved in the leveling-out of brushmarks in a paint film. Taking into account typical values of depth and width of brushmark, viscosity and surface tension of paint, Waring concludes that a paint which will not flow when subjected to a shear stress of less than two or three dynes/cm² cannot be expected to level-out brushmarks satisfactorily. If such a paint is presumed to possess a viscosity coefficient of 10 poises, it follows that the rate of shear to be anticipated for a stress of, say, one dyne/cm² greater than its yield value would be of the order of 0.1 sec^{-1}. Hence, the instrument should be sensitive to at least one percent of this rate of shear, or 0.001 sec^{-1}. Likewise, the shear stress measurement should be accurate to approximately 0.01 dyne/cm².

The Translated Cylinder Viscometer

The instrument used in the investigation herein described is based upon the principle of the translated coaxial cylinder viscometer described by Pochettino[8] and frequency referred to as the Pochettino type. Such an instrument has been employed by, among others, Traxler, Romberg, and Schweyer[5] to measure the flow properties of bitumen and by Bikermann[9] for rosin suspensions and adhesives. The latter author provides an extensive bibliography of the subject. Eirich[10] also comments on the Pochettino-type viscometer.

Figure 1 presents a diagram of the translated cylinder viscometer, which consists of a solid cylindrical bob of

* Present Address: Wright Aeronautical Division, Curtiss-Wright Corporation, Wood Ridge, New Jersey.

[1] E. C. Bingham, *Fluidity and Plasticity* (John Wiley and Sons, Inc., New York, 1922).

[2] H. Green, *Industrial Rheology and Rheological Structures* (John Wiley and Sons, Inc., New York, 1949).

[3] R. Houwink, *Second Report on Viscosity and Plasticity* (Nordemann Publishing Company, New York, 1938), p. 185.

[4] Stow, Horowitz, and Elliott, J. Colloid Sci. **4**, 321 (1949).

[5] Traxler, Romberg, and Schweyer, Ind. Eng. Chem. Anal. Ed. **14**, 340 (1942).

[6] A. K. Holliday, Trans. Faraday Soc. **43**, 630 (1947).

[7] R. K. Waring, J. Rheol. **2**, 307 (1931).

[8] A. Pochettino, Nuovo cimento **8**, 77 (1914).

[9] J. J. Bikermann, J. Colloid Sci. **3**, 75 (1948).

[10] E. R. Eirich, Proc. Intern. Congr. Holland **IV**-3 (1948).

radius r_i and length L supported coaxially within a vertical closed tube of radius r_a; the bob is immersed in the sample to be examined. A force F applied axially to the bob will cause it to move in the direction of F with a velocity v_i. This motion causes the material in the annulus between the two cylinders to be sheared, while material displaced by the bob flows through the annulus in a direction opposite to the direction of motion of the bob. A pressure difference thus appears between the ends of the bob of such sign that it opposes the force F.

An analysis of this problem is given by Smith, Ferry, and Schremp.[11] The generalized integral of the differential equation of fluid flow for this case is given by Page[12] as

$$v = Pr^2/4\eta L + \frac{A}{2\pi\eta L}\log_e r + B, \tag{1}$$

where v is the velocity at any radius r in the fluid of viscosity η and p is the pressure drop along the bob of length L. A and B are constants of integration.

The volume of fluid displaced per unit of time by the bob is equal to that flowing through the annulus. Referring to Fig. 1,

$$v_i\pi r_i^2 = \int_{r_i}^{r_a} v \cdot 2\pi r dr. \tag{2}$$

Through the use of Eq. 2 and the boundary conditions,

$$v=0 \text{ for } r=r_a; \quad v=v_i \text{ for } r=r_i.$$

Smith, Ferry, and Schremp[11] show that the shear rate dv/dr is given by

$$dv/dr = v_i \frac{r_a^2+r_i^2-2r^2}{r[(r_a^2+r_i^2)\log_e r_a/r_i - (r_a^2-r_i^2)]}. \tag{3}$$

Equation 3 solved for $r=r_i$ yields the expression for the shear rate at the surface of the bob,

$$dv/dr\big]_{r=r_i} = v_i \frac{r_a^2-r_i^2}{(r_a^2+r_i^2)\log_e r_a/r_i - (r_a^2-r_i^2)}. \tag{4}$$

The force F applied axially to the bob comprises two components

$$F = 2\pi r_i L\eta \frac{dv}{dr}\Big]_{r=r_i} + \pi r_i^2 p. \tag{5}$$

The first term on the right results from the viscous resistance of the motion of the bob, while the second term is the force required to balance the pressure drop along the bob. Upon obtaining expressions for p from

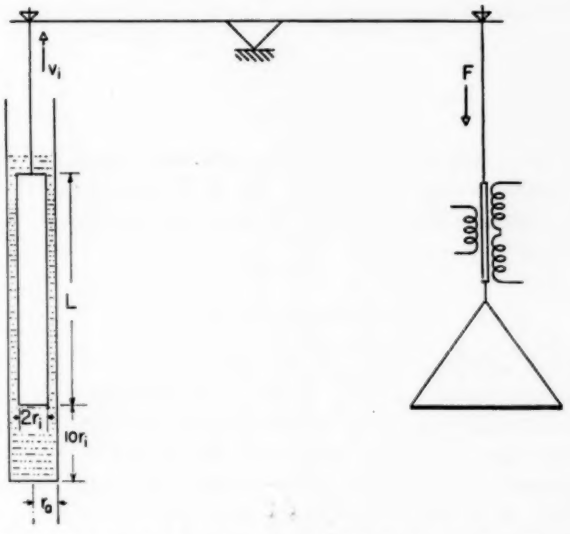


FIG. 1. Schematic diagram of translational coaxial viscometer.

Eqs. 1 and 2 and $dv/dr\big]_{r=r_i}$ from Eq. 3 in terms of v_i, r_i, and r_a, Smith *et al.* show that Eq. 5 becomes

$$F = \frac{2\pi L\eta v_i}{\log_e r_a/r_i}\left\{\frac{r_a^2-r_i^2}{(r_a^2+r_i^2)\log_e r_a/r_i - (r_a^2-r_i^2)]}+1\right\}. \tag{6}$$

Equation 6 may be rearranged to express η in terms of the variables F and v_i and the geometric constants of the instrument

$$\eta = \frac{F}{2\pi L v_i} \times\left\{\frac{[(r_a^2+r_i^2)\log_e r_a/r_i - (r_a^2-r_i^2)]\log_e r_a/r_i}{(r_a^2+r_i^2)\log_e r_a/r_i}\right\}. \tag{7}$$

This analysis for a closed-end tube indicates that the flow within the annulus is complex, resembling somewhat the conditions obtaining in a capillary tube. If the bob is rising, the liquid adjacent to the bob is rising also, and the velocity gradient (shear rate) is positive (taking v_i to be positive in the rising direction). The velocity gradient diminishes as the radial r becomes larger than r_i, vanishing at a point $r_0=(r_a^2+r_i^2)^{\frac{1}{2}}/2$. Also, here the velocity is negative, the liquid is flowing downward to replace that displaced by the motion of the bob. At the wall of the tube r_a, the shear rate is negative.

It is thus apparent that in the translational cylinder viscometer it is not possible to arrange to subject the sample to essentially a uniform stress or strain rate, as is possible in the case of a rotational viscometer. It may be pointed out that if the tube were open ended, immersed in a larger container of the liquid, the flow within the tube would no longer be complex; the viscometer would then be of the Pochettino type. The

[11] Smith, Ferry, and Schremp, J. Appl. Phys. **20**, 144 (1949).
[12] L. Page, *Introduction to Theoretical Physics* (D. Van Nostrand and Company, Inc., New York, 1947), p. 259.

expression for the translational force would now become[13]

$$F_0=\frac{2\pi\eta L v_i}{\log_e r_a/r_i},$$

whereas the equation for translational force for the closed-end tube is given by Eq. 6. Denoting this force for the closed-end tube by F_c, it is evident that

$$\frac{F_c}{F_0}=\frac{r_a^2-r_i^2}{(r_a^2+r_i^2)\log_e r_a/r_i-(r_a^2-r_i^2)}+1,$$

which will always be greater than 1 for $r_a>r_i$. That is, for the same viscosity and rate of translation v_i the closed-tube viscometer allows greater forces to be used than does a similar open-end viscometer. Since our purpose is to achieve shear stress and shear rate as small as possible, the closed-end viscometer presents an advantage which is quite considerable.†

This advantage of the closed-end tube in achieving smaller shear stress for a given force of translation than is possible with an open-end tube outweighs the disadvantage of the complex flow associated with the latter. It may be pointed out that even in the case of an open-end tube, if the sample is non-Newtonian, the velocity gradient will not be constant and the assumption of a uniform stress field within the sample might well represent a fictitious concept.

Experimental Apparatus

The translated cylinder viscometer used in this investigation consists of a brass bob suspended coaxially in a Gardner-Holt bubble viscosity tube by means of a copper wire of 0.003-in. diameter. In Fig. 1 the wire is shown suspended from one arm of an equal-arm analytical balance. The tube is rigidly clamped in a bracket affixed to the platform of the balance.

The counterbalance end of the balance beam supports the weight pan, a portion of the suspension consisting of the core of a linear variable differential transformer.‡ This transformer is an electromechanical transducer which translates the position of its free-floating core into an electrical signal which may be suitably amplified. Since the displacement of the core corresponds exactly to the displacement of the bob of the viscometer, it is thus possible to measure directly the displacement of the bob in terms of the transformer output signal.

The transformer is energized and its signal amplified by means of a bridge oscillator and amplifier described by Burr.[14] The amplifier output is a dc signal which is recorded by a dc recorder§ shown. The transformer is mounted on a micrometer screw bracket to which is attached a dial indicator. Calibration of the transformer, amplifier, and recorder in terms of core displacement is accomplished as follows: With the balance beam suspension locked, the dial indicator is set to zero near the electrical null point of the transformer by adjustment of the micrometer screw. The recorder is then electrically balanced to zero through null adjustment of the amplifier. By then moving the transformer with the micrometer screw a desired amount, noting the deflection of the recorder, the amplifier gain may be adjusted to yield a full-scale deflection of th ‹recorder to correspond to the desired core displacement. Since the amplifier is phase sensitive, the motion of the pen indicates not only the magnitude but also the direction of motion of the core in relation to the transformer.

Surrounding the fixed bracket into which the Gardner-Holt tube is clamped is a water bath through which water from a thermostated bath is pumped. Except where noted the temperature of this bath is 25°C±0.05°; a pumping rate corresponding to about two bath changes per minute suffices to eliminate temperature variations in the bath.

The viscometer tube is centered with respect to the bob by replacing the bob with a small pointed plumb bob. Inserted in the tube is a brass cylinder tapered accurately to a conical point. By adjusting the point of the plumb bob and that of the tapered cylinder to coincide, coaxial alignment is achieved to within a thousandth of an inch or so. Minor adjustments of such alignment may be made by adjusting the position of the balance beam arrest support.

Calibration of the Translational Viscometer

The calibration of the transformer, amplifier, and recorder was carried out as described above; experience indicated that a full-scale deflection of the recorder, for a displacement of 0.004 in. of the transformer core, represented the greatest sensitivity which could be employed without encountering undue variations occasioned by microseisms of the building or by the noise level and inherent stability of the amplifier. At this level of amplification (2500 to 1) the electrical system was completely stable for periods of at least two hours.

The chainomatic loading device was readily calibrated and found to correspond to 0.138 milligram per least division. Loads in excess of those delivered by the chain were applied by addition of chip weights to the balance pan.

A typical run was carried out as follows: Having placed the sample tube and sample, with bob suspended

[13] Guy Barr, *A Monograph of Viscometry* (Oxford University Press, London, 1931).

† For example, for the viscometer described herein $r_a=0.536$ cm; $r_i=0.398$ cm, and $F_c/F_0=33.7$, that is, for the open-end tube the weight units would have to be reduced by a factor of more than 1/30 to obtain equal shear stress sensitivity. Taking $dv/dr=V_i(r_a-r_i)$ as the shear rate for the open-end tube, and Eq. (4) for the closed-end tube, the shear rate in the closed tube is found to be about 3½ times that in the open-end tube. Thus the advantage in shear stress sensitivity of the closed-end tube is nearly tenfold that of the open-end tube at constant shear rate.

‡ Manufactured by Schaevitz Engineering Company, Camden, New Jersey.

[14] C. W. Burr, Electronics **22**, 101 (1949).

§ Manufactured by Leeds and Northrup Company, Philadelphia, Pennsylvania.

in it, in the fixed bracket, a period of 15 to 20 minutes was allowed to establish temperature equilibrium. During this period the transformer calibration was checked.

The balance arrest was then released, and weights were adjusted so as to cause the bob to fall at a suitable rate. The recorder was brought on scale by adjustment of the micrometer stage carrying the transformer. The motion of the bob was recorded for a sufficient period of time to produce an essentially straight displacement-time curve. For viscous liquids the curve became linear almost instantaneously; the slope of this curve represents, of course, a particular rate of shear which is taken to be *negative* when the bob is *falling*.

Next, weight was added to the pan such that the rate of shear became suitably smaller, again recording displacement *vs* time to achieve constant velocity of bob translation.

In this manner weight was added stepwise so that the bob could be observed to stop moving altogether and, in turn, to begin to rise at increasing velocity as the weight was increased. The velocity associated with *rising* of the bob is taken to be *positive*. The corresponding weight, in excess of that for which the bob ceased to move, is taken as positive weight.

To establish the experimental values of the constant in Eq. 7, the flow curves were observed for two samples of oil, the viscosity of which had been determined by the National Bureau of Standards. The data are shown graphically in Fig. 2, where the force in chainomatic units (1.000 unit = 27.6 milligrams) is plotted against the rate of translation of the bob (1 chart division per minute = 10^{-4} inch per minute). The dimensions of the viscometer are as follows:

Radius of tube, r_a: 0.536±0.0005 cm.
Radius of bob, r_i: 0.398±0.0005 cm.
Length of bob, L: 5.08 ±0.002 cm.

The expression on the right of Eq. 7 will be observed to consist of the variables F and v_i and the constants $2\pi L$ and the function of r_a and r_i within brackets. Hence Eq. 7 may be written

$$\eta = k\frac{F}{v_i} = k\frac{G}{D},$$

where G represents the applied force in chainomatic units, D represents the rate of translation of the bob in chart divisions per minute, and k includes appropriate factors to convert these variables to units of dynes and cm/sec, respectively. That is, for the dimensions given above

$$k = 1.71\times10^{3}.$$

The experimental values of k from the data in Fig. 2 are

$$k_N = 2.23\times10^{3} \quad \text{and} \quad k_M = 2.15\times10^{3}$$

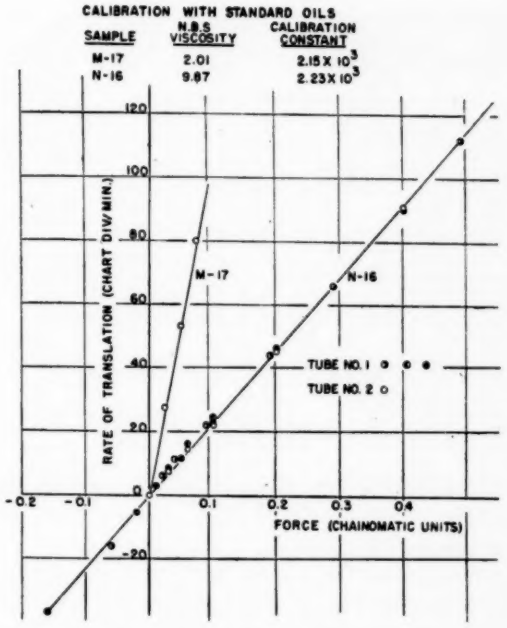


FIG. 2. Experimental calibration of translational viscometer.

for the known viscosity samples N–16 and M–17, respectively.

The discrepancy between the theoretical and observed calibration constants may be attributed to the failure to include in the analysis recognition of the forces involved in flow of the fluid around the ends of the bob. These "end effects" evidently amount to about 35 percent of the total forces; hence, the applied force must be corrected by a factor $k/k_N = 0.763$ in computing the actual shear stress.

The maximum shear rate from Eq. 4 is thus

$$dv/dr\big]_{r=r_i} = 4.82\times10^{-2}CD \text{ sec}^{-1},$$

where C is the bob deflection for full-scale pen deflection and D is the observed rate of deflection in units of chart divisions per minute.

The maximum shear stress may be computed from the Newtonian equation

$$\tau_{r=r_i} = \eta\frac{dv}{dr}\Big]_{r=r_i} = 3.91\times10^{-2}W \text{ dyne/cm}^2,$$

where η and $dv/dr]_{r=r_i}$ are experimentally determined and W is the translating force, milligrams. This constant includes the end effect correction noted above.

This method of determining the balance point, where the bob velocity is zero, obviates the necessity for bouyancy corrections for the weight of the bob.

DISCUSSION

Anomalous Flow of Complex Systems

The calibration described above indicates that the translational viscometer serves to measure the flow of

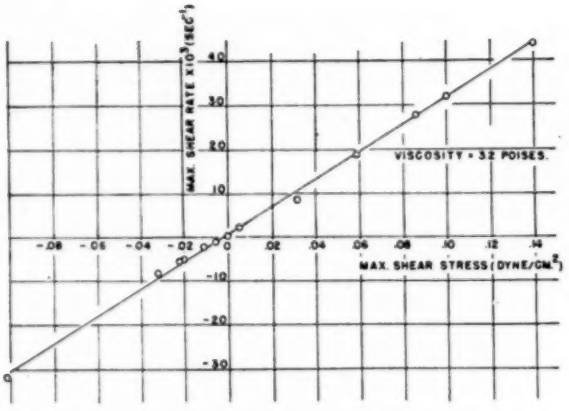


FIG. 3. Flow curve of Type III Latex (60 percent solids).

Newtonian liquids with a sensitivity of approximately 5×10^{-3} dyne/cm² maximum shear stress and 1×10^{-4} sec^{-1} maximum shear rate.

The flow curve for a high solids Latex (60 percent solids) shown in Fig. 3. This suspension, despite the large amount of solids present, appears to be completely Newtonian at all rates of shear with no evidence of a yield value. This conclusion was also reached by Krieger and Maron[15] though for considerably higher rate of shear.

It is of interest to note that spherical particles, even under such conditions of crowding, do not develop mutual stress fields greater than 0.005 dyne/cm².

A suspension of a synthetic rubber (Chemigum III) of approximately 20 percent solids in methyl-ethyl ketone exhibited the flow properties shown in Fig. 4. In this case the fluid is no longer a Newtonian liquid, and the following modification in treatment of the data was necessary.

The determination of the flow curve is initiated with an arbitrary negative force (falling bob), and upon adding weight (decreasing stress) the flow curve exhibits a singularity or flat portion for zero rate of shear. Somewhere within this "flat" singularity the bob is obviously counterbalanced, but to know the exact point at which counterbalancing occurs (i.e., where shear stress is zero), except for a Newtonian liquid, requires an extremely precise knowledge of the density and volume of the bob and the density of the sample. Since such precision is not practical the alternative is to assume that, as the stress is diminished and the shear velocity approaches zero, an elastic structure is formed at the negative shear stress corresponding to the yield value. This structure is further assumed to persist until the shear stress has passed through zero and increased again to a positive value corresponding to the yield value. The balance point, or shear stress origin, thus lies at the center of the zero-flow singularity of the flow curve.

For a Newtonian fluid, the stress distribution in the sample may be computed from Eq. 4 and the viscosity of the sample. If now this viscous fluid is replaced by an elastic solid the stresses will be identical, for constant force F, as has been shown by one of the authors.[16] In this elastic analog of the viscous case, elastic shear displacements replace viscous shear velocities. Hence, the calibration of the viscometer, by means of standard viscosity samples, is correct for the region of shear stress less than the yield value. Therefore, the maximum shear stress, at which flow is incipient, does represent the actual yield value.

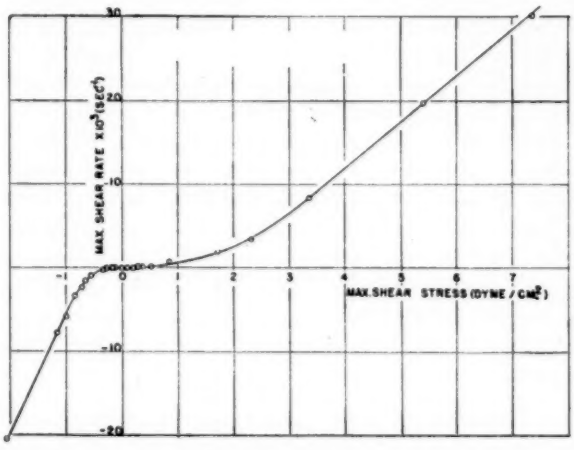


FIG. 5. Flow curve of a natural gum suspension (gum tragacanth, 1.60 percent solids in water), yield value = 0.4 dyne/cm².

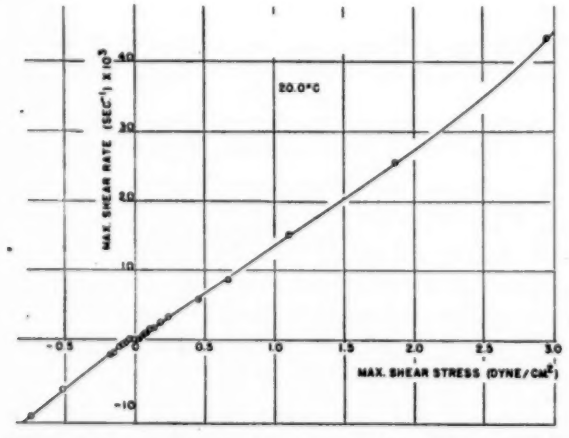


FIG. 4. Flow curve of rubber suspension (Chemigum III in methyl-ethyl ketone, 20 percent solids) at 20°C, yield value = 0.05 dyne/cm².

Once flow has commenced, the analysis is less straightforward. The non-Newtonian fluid is subjected to a complex stress field, and since only the integral of the unknown flow function is observed experimentally, the data cannot be reduced to absolute parameters. The portions of the flow curve where shear rate is finite thus cannot be treated as absolute flow curves. An empirical correlation is indicated later in this paper between the

[15] I. M. Krieger and S. H. Maron, J. Colloid Sci. 6, 528 (1951).

[16] J. T. Bergen, Proc. Soc. Exptl. Stress Anal. Semiann. Rept. 9, (II) 13 (1952).

flow curve for finite shear rate and the behavior of non-Newtonian liquids under conditions which are of technological interest.

The maximum shear stress for which flow is incipient, in Fig. 4, is 0.05 dyne/cm², this being the yield value f_0 by our conventions. The linearity of the flow curve above f_0 strongly suggests a Bingham material, although anomalous behavior sets in at higher stresses. Anomalous flow for concentrated GRS suspensions is reported by Bestul and Belcher[17] for shear rate upwards of 100 sec^{-1}, in agreement with the curves of Fig. 4.

Very marked anomalous flow is evident in Fig. 5 for a gum tragacanth suspension. The yield value appears to be 0.4 dyne/cm². Hastewell and Riston,[18] for a gum tragacanth suspension of 1.83 percent solids, postulate a viscous element of 52 poises coupled with a Maxwell, or retarded elastic element, of 45 poises. The deformation-time relation for such a model is asymptotic and the data of Fig. 5 may represent a condition of psuedo-steady flow at low shear stress. At higher stress steady flow would be established more rapidly and the flow curve should tend to become linear, in agreement with the data of Hastewell and Riston. Higginbotham,[19] however, reports anomalous flow for an unspecified gum tragacanth suspension at high shear stress.

A retarded elastic element likewise would lead to asymmetry of the flow curve about the origin, as is evident in Fig. 5.

TABLE I. Yield values for several types of household paints, as measured with the translational viscometer.

Brand	Yield value dyne/cm²	Leveling	Sag	Yield value after Fischer (see reference 7)*
		Exterior white paint		
A	1.0	Fair	Slight	110–170 dyne/cm²
B	5.1	Poor	Nil	
C	0.9	Good	Marked	
		Interior flat wall paint (white)		
A	4.7	Fair	Nil	40– 50 dyne/cm²
B	2.9	Fair	Nil	
C	0.0	Good	Marked	
		Interior semi-matte enamel (white)		
A	1.0	Good	Marked	65–170 dyne/cm²
B	5.8	Very Poor	Nil	
C	6.6	Fair	Marked	

* Figures of Fischer do not refer to particular brands.

The Yield Value of Household Paints

The results of a study of three types of household paints are given in Table I. The flow curves for three brands of one of these types (outside white) are given in Fig. 6. The corresponding flow curves for the other types were quite similar to those shown in Fig. 6. Evaluation of the leveling and sag properties of each of these paints was carried out subjectively by brushing the sample onto a hiding-power card, allowing the sample to dry with the card standing vertically. After two days' drying time the brush-outs were examined

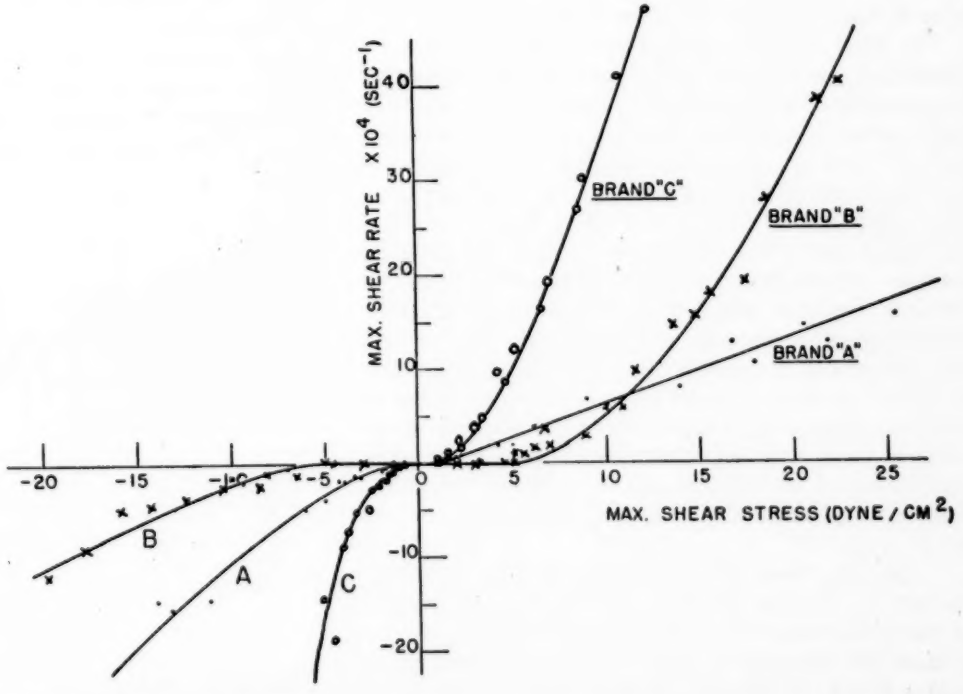


FIG. 6. Flow curve of three different brands of exterior white paint. See Table I for yield values.

[17] A. B. Bestul and H. V. Belcher, J. Colloid Sci. 5, 303 (1950).
[18] L. J. Hastewell and L. J. U. Riston, J. Sci. Instr. 29, 20 (1952).
[19] R. S. Higginbotham, J. Sci. Instr. 27, 139 (1950).

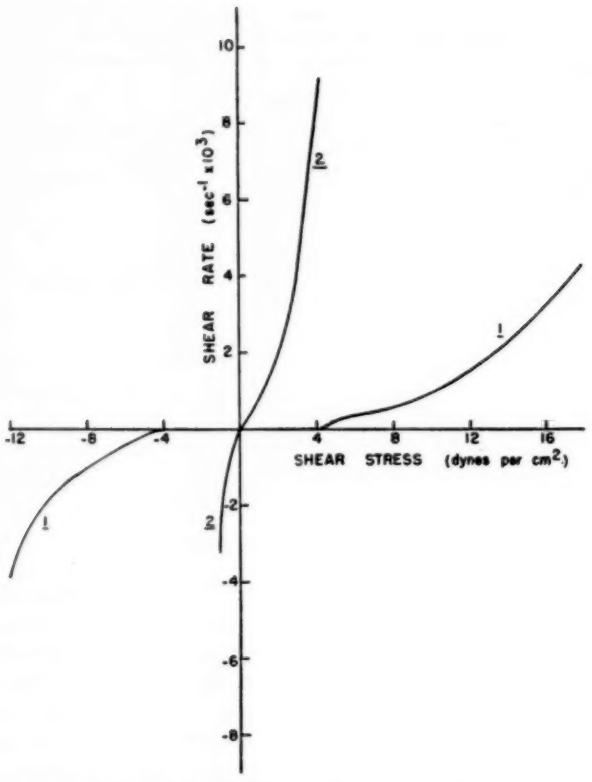


FIG. 7. Flow curve of Organosol coatings. Sample 1 leveled poorly, yield value = 4 dyne/cm². Sample 2 = sample 1 + 1 percent of a "flow-out" agent and leveled very well, yield value = nil.

visually for disappearance of brushmarks and for sagging or running of the film from a heavy "bead."

The yield value in Table I is the greatest stress for which flow was incipient in the translational viscometer. A comparison of these data indicates that paints with yield values of the order of one dyne/cm² or less level but will tend to sag or form curtains. Paints which do not level well exhibit a yield value of 5–6 dynes/cm². Two samples, Flat Wall A and Semi-Matte C showed relatively high yield values but good leveling. These samples differed from the others in requiring about 10 minutes to exhibit steady flow, upon a change of stress, where the other samples reached steady flow in two minutes or less. Evidently the longer thixotropic time scale allowed these samples to level before the onset of structure became significant, while the observed yield value is that of the equilibrium structure. A complete description of leveling properties evidently must include a consideration of the thixotropic properties of the paint.

These data, however, lend support to the conclusions reached by Waring (see reference 7) on the magnitude of the yield value for brushing paints. Waring's conclusions are disputed by Fischer[20] who reported yield values shown in Table I. Using a form of rotational viscometer he obtained flow curves ranging downwards to about 1 sec^{-1}. Extrapolation of these curves after the manner suggested by Green[2] to zero rate of shear yielded a stress axis intercept which has been denoted by the latter as "yield value." Hence, the figures obtained by Fischer in this manner represent an unjustifiably great extrapolation in terms of Fig. 6. The comparison renders the procedure for establishing yield value, as followed by Green (see reference 2) and Fischer,[20] of doubtful validity.

Owing to the limitations of the theory of flow of non-Newtonian liquids for finite rate of shear, as previously described, no absolute parameters may be derived from the flow curves of these paints. However, it has been observed that, in general, those with "shallow" flow curves, for example, Brand "A" in Fig. 6 resist sagging much better than those with "steep" curves (lower relative "viscosity") as exemplified by Brand "C," although the yield values for these cases are almost identical.

Organosol Coatings

Organosol coatings may be applied industrially in the form of a relatively thick nonuniform coating. It is essential that leveling occur before the coating has set. In the course of examination of such materials, an example of the marked effect of surface active additives, in small quantities, was observed as shown in Fig. 7. Sample 1 in this figure leveled very poorly, even in a thick film with very slow solvent evaporation. Upon adding 1 percent of a "flow-out" agent, the coating leveled very satisfactorily and the flow curve of sample 2 was obtained.

Such coatings have been observed to undergo marked changes in flow properties upon aging, depending upon the plasticizer-solvent system employed. Figure 8 illustrates the course of aging for one such

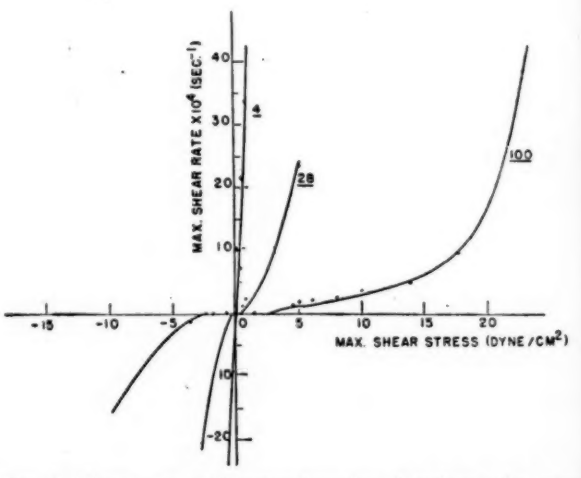


FIG. 8. Flow curve of Organosol coating. Numbers refer to age of coating, hours. At 4 hours the coating leveled very well and yield value = nil; at 28 hours leveling was fair, yield value = 0.4 dyne/cm²; at 100 hours leveling was poor, yield value = 3.8 dyne/cm².

[20] E. K. Fischer, J. Colloid Sci. 5, 271 (1950).

sample, for which the change in leveling upon aging was accompanied by a corresponding increase in yield value and viscous character of the flow curve at low rate of shear.

CONCLUSIONS

A translational concentric cylinder viscometer of the Pochettino type is described, for which an analytical balance is employed to achieve very small shearing stress and an electrical transducer to observe low rate of shear. The ultimate sensitivity of such a device is found to be 5×10^{-3} dyne/cm^2 maximum shear stress and 1×10^{-4} sec^{-1} maximum shear rate. Calibration with Newtonian liquids of known viscosity agrees with a theoretical expression for viscosity, allowing for "end" effects. The flow properties of various complex, anomalous systems are described for the region very near the shear rate-shear stress origin. With the exception of a high solids Latex suspension, which was Newtonian, the complex systems exhibited a yield value and anomalous flow in the post-yield region.

The yield values of several household paints for very small rates of shear are found to correlate with their leveling and sagging properties except where thixotropic effects become large in terms of the time required for leveling to occur. The yield values thus observed are considerably less than those reported elsewhere and do not exceed about 6 dynes/cm^2. Industrial Organosol coatings are found to possess similar characteristics. Examples are given of the effect of addition of a "flow-out" agent and of the effect of aging upon such materials.

Regular Contributed Original Research

Signal-to-Noise Ratios in Band-Pass Limiters*

W. B. DAVENPORT, JR.
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received August 26, 1952)

A general analysis is made of the relations between output signal and noise powers and input signal and noise powers for band-pass limiters having odd symmetry in their limiting characteristics. Specific results are given for the case where the limiter has an nth root characteristic, and they include the ideal symmetrical limiter (or clipper) as a limiting case. This analysis shows that, for the band-pass limiter, the output signal-fo-noise power ratio is essentially directly proportional to the input signal-to-noise power ratio for all values of the latter. This result is due to the band-pass characteristics rather than to the symmetrical limiting action.

I. INTRODUCTION

SATURATION, or limiting, may often take place in the band-pass amplifier stages of a radio receiver. Sometimes this limiting is inadvertent (as in the case of a much larger than usual input signal which overdrives an amplifier), while sometimes the band-pass stages are deliberately designed to limit (as in the case of most FM receivers). In any case, whether the limiting is inadvertent or deliberate, it is of interest to know quantitatively the action of the band-pass limiter.

The purpose of this report is to determine the relations between the output signal and noise and the input signal and noise for the case of a band-pass limiter for all values of the input signal-to-noise ratio. Previous studies have either considered the effects of noise alone as a limiter input;[1] or, when a signal-plus-noise input was considered, either the action of the limiter was studied only in combination with that of a discriminator[2-4] or results were obtained only for large values of the input signal-to-noise ratio.[5]

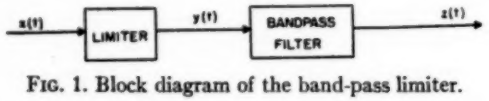


FIG. 1. Block diagram of the band-pass limiter.

The system to be considered here consists of a limiter followed by a band-pass filter, as shown in Fig. 1. The input to this system is assumed to consist of an amplitude-modulated sine wave plus a noise wave

$$x(t)\equiv P(t)\cos pt+N(t). \tag{1}$$

The input noise $N(t)$ is assumed to be Gaussian in nature and to have a narrow band spectrum centered in the vicinity of the signal carrier frequency $p/2\pi$. The spectrum of the limiter output $y(t)$ will consist of signal and noise terms centered on the angular frequencies $\pm mp$, where $m=0, 1, 2, \cdots$. The band-pass filter is assumed to have an ideal rectangular pass-band transfer characteristic which is centered on the fundamental angular frequency p. The filter pass band is assumed to be wide enough to pass all of the limiter output spectrum centered about $\pm p$, but narrow enough to reject those parts of the spectrum centered on $\pm mp$ (where $m\neq 1$).

In the analysis to follow, we will obtain expressions for the autocorrelation function at the input to the band-pass filter, as well as expressions for the signal power and noise power at the output of the band-pass filter. From these expressions we will be able to determine the relation between the output signal-to-noise power ratio and the input signal-to-noise power ratio.

II. GENERAL ANALYSIS

Let us first consider the problem of determining the autocorrelation function of the limiter output $y(t)$. This function is defined as the statistical average of $y(t)y(t+\tau)$, i.e.,

$$R_y(\tau)\equiv\langle y(t)y(t+\tau)\rangle_{Av} \tag{2}$$

and has been shown by Wiener[6] to be the Fourier transform of the spectral density.

Rice[7] and Middleton[8] have shown that if the output of a nonlinear device may be expressed as a unique function of its input

$$y(t)=g[x(t)], \tag{3}$$

and if the input to that device is an amplitude-modulated sine wave plus gaussian noise, as in Eq. (1), then the autocorrelation function of the output of the nonlinear device may be expressed as

$$R_y(\tau)=\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\frac{\epsilon_m}{k!}\langle h_{mk}(t)h_{mk}(t+\tau)\rangle_{Av}R_N{}^k(\tau)\cos mp\tau, \tag{4}$$

where the ϵ_m are the Neumann numbers

$$\begin{aligned}\epsilon_0&=1,\\ \epsilon_m&=2 \quad \text{for } m\geqslant 1,\end{aligned} \tag{5}$$

where $R_N(\tau)$ is the autocorrelation function of the input

* This work has been supported in part by the U. S. Signal Corps, the Air Materiel Command, and the U. S. Office of Naval Research.

[1] D. Middleton, J. Appl. Phys. **17**, 778 (1946).
[2] N. M. Blachman, J. Appl. Phys. **20**, 38 (1949).
[3] D. Middleton, J. Appl. Phys. **20**, 334 (1949).
[4] D. Middleton, Quart. Appl. Math. **7**, 129 (1949), and **8**, 59 (1950).
[5] D. G. Tucker, Wireless Engineer **29**, 128 (1952).
[6] N. Wiener, Acta Math. **55**, 117 (1930).
[7] S. O. Rice, Bell System Tech. J. **24**, 46 (1945).
[8] D. Middleton, Quart. Appl. Math. **5**, 445 (1948).

noise, and where the function $h_{mk}(t)$ is defined by

$$h_{mk}(t)\equiv\frac{j^{m+k}}{2\pi}\int_{-\infty}^{\infty}f(ju)e^{-[\sigma^2(N)u^2]/2}u^k J_m[uP(t)]du. \quad (6)$$

$\sigma^2(N)$ is the variance of the input noise, J_m is the mth order Bessel function of the first kind, and $f(ju)$ is the Fourier transform of the transfer characteristic of the nonlinear device

$$f(ju)\equiv\int_{-\infty}^{\infty}g(x)e^{-jux}dx. \quad (7)$$

In our present study we wish to specify that the limiter transfer characteristic $g(x)$ is a nondecreasing odd function of its argument

$$g(x)\equiv\begin{cases}g_+(x) & \text{for } x>0\\ -g_+(-x) & \text{for } x<0.\end{cases} \quad (8)$$

Because of convergence difficulties, the Fourier transform $f(ju)$ must be replaced in this case by the bilateral Laplace transform[9,10] $f(w)$:

$$f(w)=f_+(w)+f_-(w), \quad (9)$$

where w is the complex variable $v+ju$, and where $f_+(w)$ and $f_-(w)$ are the unilateral Laplace transforms of $g(x)$ in the intervals $(0, +\infty)$ and $(-\infty, 0)$, respectively. The unilateral Laplace transforms $f_+(w)$ and $f_-(w)$ have different regions of convergence in the w plane, and, as we shall see later, each may have a singularity at the origin of that plane. Because of these different regions of convergence for $f_+(w)$ and $f_-(w)$, we must in general employ two inversion integrals with separate integration contours in order to return to $g(x)$ from $f(w)$. For this reason, the single integral expression for the function $h_{mk}(t)$ must be replaced by the sum of two contour integrals in this study

$$h_{mk}(t)=\frac{1}{2\pi j}\int_{C+}f_+(w)e^{[\sigma^2(N)w^2]/2}w^k I_m[wP(t)]dw$$

$$+\frac{1}{2\pi j}\int_{C-}f_-(w)e^{[\sigma^2(N)w^2]/2}w^k I_m[wP(t)]dw, \quad (10)$$

where $C+$ is the contour along the imaginary axis of the w plane with a possible indentation to the right of the origin, and $C-$ is the contour along the imaginary axis with a possible indentation to the left of the origin.

Because of the odd character of $g(x)$, it follows that

$$f_-(w)=-f_+(-w) \quad (11)$$

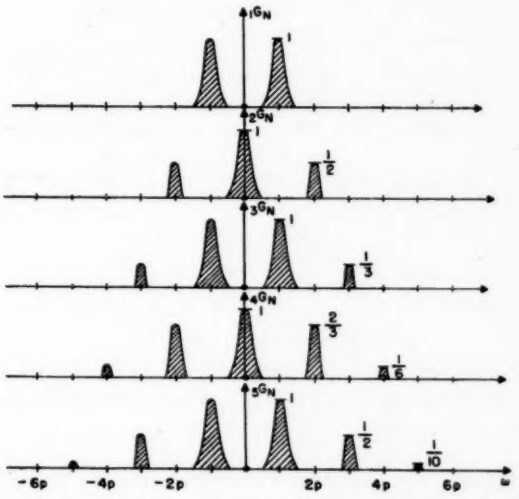


FIG. 2. Plots of the spectral densities ${}_kG_N(\omega)$, showing the relative values of the various spectral contributions.

and therefore that

$$h_{mk}(t)=\begin{cases}0 & \text{for } m+k \text{ even}\\ 2\cdot\dfrac{1}{2\pi j}\displaystyle\int_{C+}f_+(w)e^{[\sigma^2(N)w^2]/2}w^k I_m[wP(t)]dw & \text{for } m+k \text{ odd.}\end{cases} \quad (12)$$

Thus we see that, because of our assumed odd symmetry for the limiter transfer characteristic, the functions $h_{mk}(t)$ vanish whenever the sum of the indices $m+k$ is even. Using this extended definition for $h_{mk}(t)$, we may now use Eq. (4) to determine the autocorrelation function of the limiter output.

From this point on, we will for convenience assume that the input signal is unmodulated. That is, we will assume that $P(t)$ is a constant. In this case, the functions $h_{mk}(t)$ are not functions of t, and Eq. (4) for the limiter output autocorrelation function simplifies to

$$R_y(\tau)=\sum_{\substack{m=0\\(m+k\text{ odd})}}^{\infty}\sum_{k=0}^{\infty}\frac{\epsilon_m h_{mk}^2}{k!}R_N{}^k(\tau)\cos mp\tau, \quad (13)$$

where the coefficients h_{mk} are determined from Eq. (12). It is convenient to expand Eq. (13) as follows

$$R_y(\tau)=2\sum_{\substack{m=1\\(m\text{ odd})}}^{\infty}h_{m0}{}^2\cos mp\tau+\sum_{\substack{k=1\\(k\text{ odd})}}^{\infty}\frac{h_{0k}{}^2}{k!}R_N{}^k(\tau)$$

$$+2\sum_{\substack{m=1\\(m+k\text{ odd})}}^{\infty}\sum_{k=1}^{\infty}\frac{h_{mk}{}^2}{k!}R_N{}^k(\tau)\cos mp\tau. \quad (14)$$

The first set of terms (sum over m) is periodic and consists of the signal output terms representing the interaction of the input signal with itself ($S\times S$ terms). The remaining terms are the limiter output noise terms.

[9] E. C. Titchmarsh, *Fourier Integrals* (Clarendon Press, Oxford, 1948), second edition.

[10] B. van der Pol and H. Bremmer, *Operational Calculus* (Cambridge University Press, Cambridge, 1950).

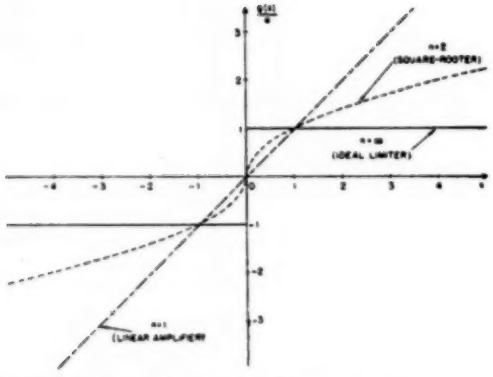


FIG. 3. Plots of the transfer characteristics of several cases of the band-pass rooter.

The second set (sum over k) represents the interaction of the input noise with itself ($N\times N$ terms), while the last set (sum over m and k) represents the interaction of the input signal and noise ($S\times N$ terms).

The spectral density of the limiter output is simply the Fourier transform of the autocorrelation function $R_y(\tau)$, i.e.,

$$G_y(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}R_y(\tau)e^{-j\omega\tau}d\tau, \tag{15}$$

where $G_y(\omega)$ is the so-called "two-sided" spectral density containing both positive and negative frequencies. The Fourier transform of Eq. (14) may be written as

$$G_y(\omega)=\sum_{\substack{m=1\\(m\text{ odd})}}^{\infty}h_{m0}^2\cdot 2\pi[\delta(\omega-mp)+\delta(\omega+mp)]+\sum_{\substack{k=1\\(k\text{ odd})}}^{\infty}\frac{h_{0k}^2}{k!}{}_kG_N(\omega)+\sum_{\substack{m=1\\(m+k\text{ odd})}}^{\infty}\sum_{k=1}^{\infty}\frac{h_{mk}^2}{k!}\times[{}_kG_N(\omega-mp)+{}_kG_N(\omega+mp)], \tag{16}$$

where $\delta(\omega)$ is the unit impulse function[10] (Dirac delta-function), and where ${}_kG_N(\omega)$ is the Fourier transform of $R_N{}^k(\omega)$. Successive applications of the convolution theorem[10] shows that ${}_kG_N(\omega)$ may be expressed as the $(k-1)$-fold convolution of $G_N(\omega)$ with itself:

$${}_kG_N(\omega)=\frac{1}{(2\pi)^{k-1}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}G_N(\omega_{k-1})\times G_N(\omega_{k-2}-\omega_{k-1})\cdots G_N(\omega-\omega_1)d\omega_{k-1}\cdots d\omega_1, \tag{17}$$

where $G_N(\omega)$ is the spectral density of the input noise. One may obtain an idea of the form of the various ${}_kG_N$ by specifying the form of the spectral density of the input noise, and then using Eq. (17) to construct plots of the ${}_kG_N$ up to any desired value of k. Plots were obtained in this manner for several values of k, and are shown in Fig. 2, along with the relative heights of the various spectral contributions. It follows from Eq. (17) that each of the spectral regions in a particular ${}_kG_N$ has the same shape, although this shape may differ from that obtained with a different value of k.

Let us now consider the system output. Because of the band-pass filter, only those terms in Eqs. (14) and (16) which correspond to energy in the vicinity of $\pm p$ will appear at the system output. The output signal autocorrelation function is therefore given by

$$R_{S0}(\tau)=2h_{10}^2\cos p\tau \tag{18}$$

and the signal output spectral density is

$$G_{S0}(\omega)=h_{10}^2\cdot 2\pi[\delta(\omega-p)+\delta(\omega+p)]. \tag{19}$$

The output signal power may be obtained by setting τ equal to zero in Eq. (18)

$$S_0\equiv R_{S0}(0)=2h_{10}^2. \tag{20}$$

The noise terms in the filter output may also be obtained by picking out those terms in Eqs. (14) and (16) which contribute only to the spectral region in the vicinity of $\pm p$. Examination of Eqs. (14) and (16) shows that only those ($N\times N$) terms corresponding to odd values of k appear. From Fig. 2, we then see that all of these terms contribute to the noise output from the band-pass filter. We may determine the filter output noise power due to these terms by setting τ equal to zero in the appropriate terms in Eq. (14), and then multiplying each term in the resultant series by a factor representing the fraction of that term that appears at the filter output. From such a process, we find that the filter output noise power caused by the interaction of the input noise with itself is given by

$$N_{0(N\times N)}=1\cdot\frac{h_{01}^2}{1!}R_N(0)+\frac{3}{4}\cdot\frac{h_{03}^2}{3!}R_N{}^3(0)+\frac{5}{8}\cdot\frac{h_{05}^2}{5!}R_N{}^5(0)+\frac{35}{64}\cdot\frac{h_{07}^2}{7!}R_N{}^7(0)+\cdots. \tag{21}$$

Plots of the spectra of the various ($S\times N$) noise terms at the limiter output may easily be constructed from the plots of Fig. 2 by translating each plot in that figure by an amount $\pm mp$ along the ω axis. A study of Fig. 2 shows that there is a maximum value of the shift mp which will allow a given term to contribute to the spectral region in the vicinity of p. From such a study, we see that for a given k, the only significant values of m are those in the range $(1, k+1)$, such that $(m+k)$ is odd. Therefore, the only terms in Eq. (14) that can contribute to the filter output noise are those given by

$$R'_{y(S\times N)}(\tau)=2\sum_{\substack{m=1\\(m+k\text{ odd})}}^{k+1}\sum_{k=1}^{\infty}\frac{h_{mk}^2}{k!}R_N{}^k(\tau)\cos mp\tau. \tag{22}$$

We may then find the filter output noise power due to the ($S\times N$) terms by setting τ equal to zero in Eq. (22) and multiplying each term by an appropriate factor

corresponding to the fraction of that term appearing at the filter output. The result is

$$N_{0(S\times N)}=\frac{1}{2}\cdot 2\frac{h_{21}^2}{1!}R_N(0)+\left(\frac{3}{4}\cdot 2\frac{h_{12}^2}{2!}+\frac{1}{4}\cdot 2\frac{h_{32}^2}{2!}\right)R_N^2(0)$$
$$+\left(\frac{1}{2}\cdot 2\frac{h_{23}^2}{3!}+\frac{1}{8}\cdot 2\frac{h_{43}^2}{3!}\right)R_N^3(0)$$
$$+\left(\frac{5}{8}\cdot 2\frac{h_{14}^2}{4!}+\frac{5}{16}\cdot 2\frac{h_{34}^2}{4!}+\frac{1}{16}\cdot 2\frac{h_{54}^2}{4!}\right)R_N^4(0)+\cdots. \quad (23)$$

The total filter output noise power is then given by the sum of the $(N\times N)$ terms and the $(S\times N)$ terms, and the filter output signal-to-noise power ratio is defined as the ratio of the filter output signal power S_0 to the total filter output noise power N_0. In order to proceed further, we will have to assume a specific form for the limiter transfer characteristic $g(x)$. From this transfer characteristic, we may then determine the coefficients h_{mk} and substitute the result in the above expressions.

III. THE BAND-PASS ROOTER

A. General Results

A convenient form of limiter transfer characteristic is that where the limiter output is proportional to the nth root of its input. We then have

$$g_+(x)\equiv\begin{cases}\alpha x^{1/n} & \text{for } x\geq 0\\ 0 & \text{for } x<0,\end{cases} \quad (24)$$

where α is a scaling constant. Plots of $g(x)$ for several values of n are given in Fig. 3. The case $(n=1)$ corresponds to the linear amplifier, while the case $(n=\infty)$ corresponds to the ideal symmetrical limiter.

The Laplace transform $f_+(w)$ of this transfer characteristic is readily determined to be

$$f_+(w)=\alpha\frac{\Gamma\left(1+\frac{1}{n}\right)}{w^{1+(1/n)}}, \quad (25)$$

where Γ is the usual gamma-function. In general, this transform has a branch point at the origin of the w plane, and the contour $C+$ must consequently have an indentation to the right of the origin.

The coefficient h_{mk} is given by substitution of Eq. (25) in Eq. (12) so that

$$h_{mk}=\begin{cases}0 & \text{for } m+k \text{ even}\\ 2\alpha\Gamma\left(1+\frac{1}{n}\right)\cdot\frac{1}{2\pi j}\int_{C+}e^{[\sigma^2(N)w^2]/2} & \\ \quad\times w^{k-1-(1/n)}I_m(Pw)dw & \text{for } m+k \text{ odd.}\end{cases} \quad (26)$$

This contour integral is essentially the same as that required in the study of the νth law detector (if $1/n$ is replaced by ν). Paralleling the evaluation of the corresponding integral for the νth law detector,[6,7] one can readily determine that

$$h_{mk}=\begin{cases}0 & \text{for } m+k \text{ even}\\ \dfrac{\alpha\Gamma\left(1+\frac{1}{n}\right)\left[\frac{P^2}{2\sigma^2(N)}\right]^{m/2}{}_1F_1\left[\frac{m+k-(1/n)}{2};m+1;-\frac{P^2}{2\sigma^2(N)}\right]}{\left[\frac{\sigma^2(N)}{2}\right]^{[k-(1/n)]/2}\Gamma(m+1)\Gamma\left[1-\frac{m+k-(1/n)}{2}\right]} & \text{for } m+k \text{ odd,}\end{cases} \quad (27)$$

where ${}_1F_1$ is the confluent hypergeometric function:[11]

$${}_1F_1[a;c;-z]=1-\frac{a}{c}\frac{z}{1!}+\frac{a(a+1)}{c(c+1)}\frac{z^2}{2!}-\cdots. \quad (28)$$

Reference to Sec. II shows that the coefficient h_{mk} always occurs in combination with $R_N(\tau)$ in the form $h_{mk}^2R_N^k(\tau)$. Now

$$R_N^k(0)=\sigma^{2k}(N), \quad (29)$$

as the input noise has a zero mean. Then by using this result, and the fact that the input signal-to-noise power ratio is given by

$$\left(\frac{S}{N}\right)_i\equiv\frac{S_i}{N_i}=\frac{P^2/2}{\sigma^2(N)}, \quad (30)$$

[11] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949).

we may write

$$h_{mk}{}^2R_N{}^k(0)=\begin{cases}0 & \text{for } m+k \text{ even}\\ \dfrac{\alpha^2 2^{k-(1/n)}\Gamma^2\left(1+\dfrac{1}{n}\right)\sigma^{2/n}(N)\left(\dfrac{S}{N}\right)_i^m {}_1F_1{}^2\left[\dfrac{m+k-(1/n)}{2};m+1;-\left(\dfrac{S}{N}\right)_i\right]}{\Gamma^2(m+1)\Gamma^2\left[1-\dfrac{m+k-(1/n)}{2}\right]} & \text{for } m+k \text{ odd.}\end{cases} \tag{31}$$

This result may now be substituted in expressions (20), (21), and (23) in order to obtain the output signal and noise powers.

A partial check on our results may be obtained by considering the case ($n=1$). When we substitute ($n=1$) into Eq. (31), we find that the only nonvanishing coefficients are h_{10} and h_{01}, and that the output signal and noise powers are simply α^2 times the corresponding input powers. These results are comforting in view of the fact that our limiter here is a linear amplifier with a power gain α^2.

B. The Ideal Symmetrical Limiter

As may be seen from Fig. 3, the ideal symmetrical limiter may be represented as a limiting case (for $n=\infty$) of the rooter. Substitution of this limiting value of n in Eq. (31) gives

$$h_{mk}{}^2R_N{}^k(0)=\begin{cases}0 & \text{for } m+k \text{ even}\\ \dfrac{\alpha^2 2^k\left(\dfrac{S}{N}\right)_i^m {}_1F_1{}^2\left[\dfrac{m+k}{2};m+1;-\left(\dfrac{S}{N}\right)_i\right]}{\Gamma^2(m+1)\Gamma^2\left(1-\dfrac{m+k}{2}\right)} & \text{for } m+k \text{ odd.}\end{cases} \tag{32}$$

If we use this result in our previously obtained expression (20) for the output signal power, we obtain

$$S_0=\frac{2\alpha^2}{\pi}\cdot\left(\frac{S}{N}\right)_i {}_1F_1{}^2\left[\frac{1}{2};2;-\left(\frac{S}{N}\right)_i\right]. \tag{33}$$

This output signal power has been plotted in Fig. 4 as a function of the input signal-to-noise power ratio $(S/N)_i$.

An expression for the output noise power may be obtained by using Eq. (32) in our previously determined expressions (21) and (23). The result is

$$N_0=\frac{2\alpha^2}{\pi}\left\{\begin{aligned}&{}_1F_1{}^2\left[\frac{1}{2};1;-\left(\frac{S}{N}\right)_i\right]+\frac{1}{8}{}_1F_1{}^2\left[\frac{3}{2};1;-\left(\frac{S}{N}\right)_i\right]\\ &+\frac{3}{64}{}_1F_1{}^2\left[\frac{5}{2};1;-\left(\frac{S}{N}\right)_i\right]+\frac{75}{3072}{}_1F_1{}^2\left[\frac{7}{2};1;-\left(\frac{S}{N}\right)_i\right]+\cdots\\ &+\frac{1}{16}\left(\frac{S}{N}\right)_i^2{}_1F_1{}^2\left[\frac{3}{2};3;-\left(\frac{S}{N}\right)_i\right]+\frac{3}{8}\left(\frac{S}{N}\right)_i{}_1F_1{}^2\left[\frac{3}{2};2;-\left(\frac{S}{N}\right)_i\right]\\ &+\frac{1}{128}\left(\frac{S}{N}\right)_i^3{}_1F_1{}^2\left[\frac{5}{2};4;-\left(\frac{S}{N}\right)_i\right]+\frac{3}{32}\left(\frac{S}{N}\right)_i^2{}_1F_1{}^2\left[\frac{5}{2};3;-\left(\frac{S}{N}\right)_i\right]\\ &+\frac{25}{24\,576}\left(\frac{S}{N}\right)_i^4{}_1F_1{}^2\left[\frac{7}{2};5;-\left(\frac{S}{N}\right)_i\right]+\cdots\end{aligned}\right\}. \tag{34}$$

A plot of this output noise power is given in Fig. 4 as a function of $(S/N)_i$.

Even though we have in Fig. 4 plots of S_0 and N_0 as functions of $(S/N)_i$, it is desirable also to obtain approximate expressions for these powers which are valid in the regions of very large (or very small) values of the input signal-to-noise power ratio.

From the power series expansion about the origin of

the confluent hypergeometric function ${}_1F_1[a;c;-z]$ (Eq. 28), we see that

$$ {}_1F_1[a;c;-z] \to 1 \quad \text{as } z \to 0. \tag{35}$$

Using the limiting result (35) in Eq. (33), we obtain for the output signal power

$$ S_0 \approx \frac{2\alpha^2}{\pi}\left(\frac{S}{N}\right)_i \quad \text{for } \left(\frac{S}{N}\right)_i \to 0. \tag{36}$$

Referring to the numerical calculations used to obtain Fig. 4, we see that the error involved in Eq. (36) is less than ten percent when values of $(S/N)_i$ are less than about two-tenths.

By using the limiting result (35) in Eq. (34), we obtain for the output noise power

$$ N_0 \approx \frac{2\alpha^2}{\pi}\frac{4}{\pi} \quad \text{for } \left(\frac{S}{N}\right)_i \to 0. \tag{37}$$

A study of Eq. (34) shows that in the region of small $(S/N)_i$, the dominant output noise is that due to direct feedthrough of the noise input to the limiter.

From Eqs. (36) and (37) we obtain

$$ \left(\frac{S}{N}\right)_0 \approx \frac{\pi}{4}\left(\frac{S}{N}\right)_i \quad \text{for } \left(\frac{S}{N}\right)_i \to 0. \tag{38}$$

Thus we see that, for the ideal, symmetrical, band-pass limiter, the output signal-to-noise power ratio is directly proportional to the input signal-to-noise power ratio in the region of very small $(S/N)_i$. This result differs radically from the familiar square-law behavior of detectors[7,8] in the region of small $(S/N)_i$. The present result is due primarily to the fact that the output terms of interest in the present case are those in the vicinity of the input frequencies, while in the detector case, the output terms of interest are those in the vicinity of zero frequency.

From the series expansion about infinity[11] of the confluent hypergeometric function ${}_1F_1[a;c;-z]$, we see

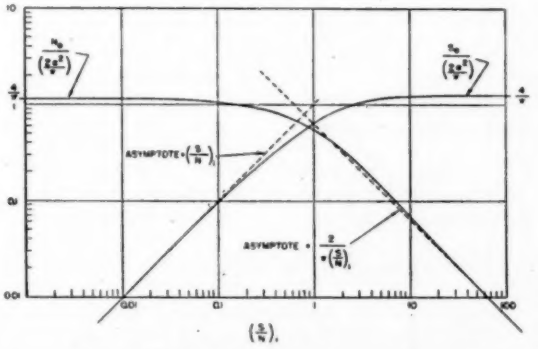


FIG. 4. Output signal power S_0 and output noise power N_0 plotted as functions of the input signal-to-noise power ratio $(S/N)_i$ for the case of the ideal, symmetrical, band-pass limiter.

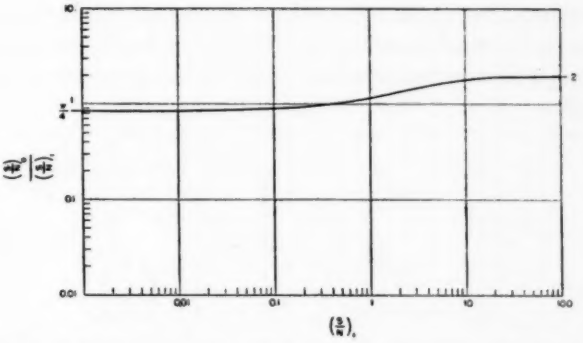


FIG. 5. The ratio of output signal-to-noise power ratio to the input signal-to-noise power ratio as a function of the latter, for the case of the ideal, symmetrical, band-pass limiter.

that

$$ {}_1F_1[a;c;-z] \approx \frac{\Gamma(c)}{\Gamma(c-a)z^a} \quad \text{for } z \to \infty. \tag{39}$$

If we now use the limiting result (39) in Eq. (33), we obtain for the output signal power

$$ S_0 \approx \frac{2\alpha^2}{\pi}\left(\frac{4}{\pi}\right) \quad \text{for } \left(\frac{S}{N}\right)_i \to \infty. \tag{40}$$

Application of the limiting expression (39) to Eq. (34) gives for the output noise power

$$ N_0 \approx \frac{2\alpha^2}{\pi} \cdot \frac{2}{\pi(S/N)_i} \quad \text{for } \left(\frac{S}{N}\right)_i \to \infty. \tag{41}$$

A study of Eq. (34) shows that the dominant noise output terms in this case are (a) the direct feed-through noise term, and (b) the noise resulting from the interaction of the input noise with the second harmonic of the input sine-wave signal.

The output signal-to-noise power ratio becomes

$$ \left(\frac{S}{N}\right)_0 \approx 2\left(\frac{S}{N}\right)_i \quad \text{for } \left(\frac{S}{N}\right)_i \to \infty. \tag{42}$$

Thus, for very large values of the input signal-to-noise power ratio, we find that the output signal-to-noise power ratio is twice that of the input. This is the result obtained by Tucker.[5]

From Eqs. (38) and (42), we see that the output signal-to-noise power ratio is directly proportional to $(S/N)_i$ in both the large and small value regions of $(S/N)_i$. In order to determine the behavior of $(S/N)_0$ for intermediate values of $(S/N)_i$, the ratio of $(S/N)_0$ to $(S/N)_i$ was calculated from the data used to plot Fig. 4, and the results were plotted to form Fig. 5. From this plot, we see that in the case of the ideal, symmetrical, band-pass limiter, the output signal-to-noise power ratio is essentially linearly proportional to the input signal-to-noise power ratio for all values of the

latter. A review of this analysis shows that this behavior is due primarily to the action of the band-pass filter following the limiter rather than to the symmetry of the limiter.

C. The Square-Rooter

Let us now consider the case of the square-rooter: Substitution of $(n=2)$ in the general expression, Eq. (31), gives

$$h_{mk}^2 R_N^k(0)=\begin{cases}0 & \text{for } m+k \text{ even}\\ \dfrac{\alpha^2\pi}{4}\dfrac{\sigma(N)}{\sqrt{2}}\left(\dfrac{S}{N}\right)_i^m \dfrac{2^k\,{}_1F_1^2\left[\dfrac{m+k-\frac{1}{2}}{2};m+1;-\left(\dfrac{S}{N}\right)_i\right]}{\Gamma^2(m+1)\Gamma^2\left(1-\dfrac{m+k-\frac{1}{2}}{2}\right)} & \text{for } m+k \text{ odd}\end{cases} \tag{43}$$

for the case of the square-rooter.

Substitution of Eq. (43) into Eq. (20) gives

$$S_0=\frac{\alpha^2}{4\pi}\Gamma^2\left(\frac{1}{4}\right)\frac{\sigma(N)}{\sqrt{2}}\left(\frac{S}{N}\right)_i {}_1F_1^2\left[\frac{1}{4};2;-\left(\frac{S}{N}\right)_i\right] \tag{44}$$

as an expression for the output signal power from the band-pass square-rooter.

Substitution of Eq. (43) into Eqs. (21) and (23) gives

$$N_0=\frac{\alpha^2}{4\pi}\Gamma^2\left(\frac{1}{4}\right)\frac{\sigma(N)}{\sqrt{2}}\left\{\begin{aligned}&{}_1F_1^2\left[\frac{1}{4};1;-\left(\frac{S}{N}\right)_i\right]+\frac{1}{32}\,{}_1F_1^2\left[\frac{5}{4};1;-\left(\frac{S}{N}\right)_i\right]\\&+\frac{25}{3072}\,{}_1F_1^2\left[\frac{9}{4};1;-\left(\frac{S}{N}\right)_i\right]+\cdots\\&+\frac{1}{64}\left(\frac{S}{N}\right)_i^2 {}_1F_1^2\left[\frac{5}{4};3;-\left(\frac{S}{N}\right)_i\right]\\&+\frac{3}{32}\left(\frac{S}{N}\right)_i {}_1F_1^2\left[\frac{5}{4};2;-\left(\frac{S}{N}\right)_i\right]\\&+\frac{25}{4608}\left(\frac{S}{N}\right)_i^3 {}_1F_1^2\left[\frac{9}{4};4;-\left(\frac{S}{N}\right)_i\right]\\&+\frac{25}{1536}\left(\frac{S}{N}\right)_i^2 {}_1F_1^2\left[\frac{9}{4};3;-\left(\frac{S}{N}\right)_i\right]\\&+\frac{125}{524\,288}\left(\frac{S}{N}\right)_i^4 {}_1F_1^2\left[\frac{13}{4};5;-\left(\frac{S}{N}\right)_i\right]+\cdots\end{aligned}\right\} \tag{45}$$

as an expression for the output noise power from the band-pass square-rooter.

Let us now obtain limiting expressions for the output signal and noise powers valid in the region of large (or small) values of input signal-to-noise power ratio. Substitution of the small $(S/N)_i$ limiting expression (35) for the confluent hypergeometric function in Eq. (44) gives for the output signal power

$$S_0\approx\frac{\alpha^2}{4\pi}\Gamma^2\left(\frac{1}{4}\right)\frac{\sigma(N)}{\sqrt{2}}\left(\frac{S}{N}\right)_i \quad \text{for } \left(\frac{S}{N}\right)_i\to 0. \tag{46}$$

Substitution of the expression (35) in Eq. (45) gives for the output noise power

$$N_0\approx\frac{\alpha^2}{4\pi}\Gamma^2\left(\frac{1}{4}\right)\frac{\sigma(N)}{\sqrt{2}}(1.04) \quad \text{for } \left(\frac{S}{N}\right)_i\to 0. \tag{47}$$

From these expressions, we obtain

$$\left(\frac{S}{N}\right)_0\approx 0.96\left(\frac{S}{N}\right)_i \quad \text{for } \left(\frac{S}{N}\right)_i\to 0. \tag{48}$$

Thus, in the region of very small values of the input

signal-to-noise power ratio, the output signal-to-noise power ratio is equal to 96 percent of the input ratio. Substitution of the large $(S/N)_i$ limiting expression (39) for the confluent hypergeometric function in Eq. (44) gives for the output signal power

$$S_0 \approx \frac{\alpha^2}{4\pi}\left(\frac{8}{9\pi^2}\right)\Gamma^2\left(\frac{1}{4}\right)\frac{\sigma(N)}{\sqrt{2}}\left(\frac{S}{N}\right)_i^{\frac{1}{2}} \quad \text{for} \quad \left(\frac{S}{N}\right)_i \to \infty. \quad (49)$$

Substitution of Eq. (39) in Eq. (45) gives for the output noise power

$$N_0 \approx \frac{\alpha^2}{4\pi}\left(\frac{5}{9\pi^2}\right)\Gamma^2\left(\frac{1}{4}\right)\frac{\sigma(N)}{\sqrt{2}}\left(\frac{S}{N}\right)_i^{-(1/2)} \quad \text{for} \quad \left(\frac{S}{N}\right)_i \to \infty. \quad (50)$$

From these expressions, we obtain

$$\left(\frac{S}{N}\right)_0 \approx \frac{8}{5}\left(\frac{S}{N}\right)_i \quad \text{for} \quad \left(\frac{S}{N}\right)_i \to \infty. \quad (51)$$

Thus, in the region of very large values of the input signal-to-noise power ratio, the output signal-to-noise power ratio is again found to be directly proportional to the input ratio. We see then, as in our previous cases, that the output signal-to-noise power ratio is essentially directly proportional to the input signal-to-noise power ratio for the case of the band-pass square-rooter.

IV. CONCLUSIONS

We have presented here an analysis of the relation between the output signal-to-noise power ratio and the input signal-to-noise power ratio for the case of a band-pass limiter. This analysis shows that, for this type of system, the output signal-to-noise power ratio is essentially directly proportional to the input signal-to-noise power ratio for all values of the latter. This type of behavior is due primarily to the fact that, in the systems studied here, the system output has been so filtered that it contains only those frequency components in the immediate vicinity of the input frequencies.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Space Charge Requirements in Some Ideally Focused Electron-Optical Systems*

NELSON WAX
University of Illinois, Urbana, Illinois
(Received March 2, 1953)

A selected set of points on the trajectories of charged particles traversing an electron-optical system may be considered to be an object space, and another set an image space. Electron-optical lenses are said to be ideally focused if the transformation of the points in object space to those in image space is one to one and continuous. A necessary condition for ideal focusing is used to prove that a number of electron-optical systems require the presence of space charge in order to be ideally focused. Expressions for some ideally focused electron-optical lenses are given, in particular those for cylindrical lenses without space charge derived from spherical lenses with space charge.

I. INTRODUCTION

A VARIETY of electron-optical systems have been used to bring a set of electron paths emerging from one point, or points, to an approximate focus at some other point, or points. The focusing usually is not perfect, unless the bundles of trajectories are of infinitesimal cross section. Since the aperture of any system is finite, it is of some interest to determine whether or not finite bundles of charged particle trajectories, emerging from one set of points, can be brought to exact focus at another set of points.

The particle trajectories may be considered as effecting a transformation of one set of points into another set. If the mapping is one to one, then all trajectories emerging from one point will be brought to exact focus at another point. The study of exact focusing systems is equivalent, therefore, to the investigation of physically realizable one-to-one transformations. One may require, further, that the transformations be continuous: curves will then map into curves. Ideal focusing will be said to occur if the transformation is physically possible, one to one, and continuous.

The above formulation of ideal focusing has long been known in geometrical optics. Gabor[1,2] appears to have been the first, however, to study ideally focused electron-optical systems. He rediscovered a necessary condition for ideal focusing and concluded, erroneously, as was pointed out by Glaser,[3] that only one electron-optical system (the Maxwell Fisheye) would yield a one-to-one continuous mapping of a three-dimensional

* This work was supported in part by the U. S. Air Forces.

[1] D. Gabor, Nature, **150**, 650 (1942), also Electronic Eng. **15**, 295, 328 and 372 (1942–1943).
[2] An account of Gabor's work is given in V. E. Cosslett *Introduction to Electron Optics* (Oxford University Press, London, 1946), p. 142.
[3] W. Glaser. Nature **162**, 455 (1948).

region into a three-dimensional region. The Maxwell Fisheye contains space charge.

The use of space-charge lenses has often been suggested as a means of reducing aberrations in electron-optical systems, but the experimental difficulties encountered in forming and maintaining the appropriate space charge configurations have not been overcome in any simple or general fashion.

The purpose of this note is to investigate the necessity for space charge in some ideally focused electron-optical systems. The work of Gabor and of Glaser is generalized, using results from geometrical optics, and analytical expressions for some ideally focused systems are developed.

Definitions and assumptions are given in the remainder of this section. A necessary condition for ideal focusing is stated and some three dimensional ideal focusing lenses are treated in Sec. II. Spherically symmetric lenses are considered in Sec. III; a class of cylindrical lenses is discussed in Sec. IV.

The following definitions and assumptions hold throughout this paper. Charged particles are emitted from one surface, the source, with constant initial speed (but in different directions). These particles constitute the primary beam. The orbits of primary beam particles traverse a region in space; a selected set of points in the region forms an object space, and another selected set, an image space. If ideal focusing occurs then a point in object space can be associated with a corresponding point in image space: the two points are termed conjugate points.

A lens is a region in space where at least one of the field intensities, electric and magnetic, is not zero. Note that the boundaries of a lens need not coincide with object or image space boundaries. Physically realizable lenses are three dimensional, but if some charged particle orbits remain in a plane section of a lens then this plane will be called a plane lens. Object and image space will be two, or one dimensional regions for plane lenses.

The electric and magnetic fields are assumed to be constant in time. Relativistic effects, the space charge effects of the primary beam, and the magnetic effects of the particle motion are ignored. All scattering and collision processes are neglected.

Space charge may be present within a lens; it is assumed, however, that the transitions between regions with and without space charge are smooth. Analytically, it is required that all first derivatives of the electric field intensity exist and be continuous everywhere.

II. GENERAL CONSIDERATIONS

A necessary condition for ideal focusing exists in the theory of geometrical optics, and hence, by the known correspondence between particle trajectories and the paths of light rays, in electron-optical theory. The correspondence is stated, the necessary condition given, and some of the consequences of this condition developed here.

Consider a point, P_0, on the path, C_0, of a light ray, and a point, P_0', on a primary beam particle's trajectory, C_0'. Let the index of refraction of the light optical medium be n, and that of the electron-optical medium be n', where

$$n' \sim k\left[(V-V_0)^{\frac{1}{2}} + \left(\frac{\eta}{2}\right)^{\frac{1}{2}} \mathbf{A}\cdot\frac{d\mathbf{s}}{ds}\right]. \tag{2.1}$$

If the slopes of the light path, and of the particle's trajectory are equal at P_0 and P_0', and if $n=n'$ along the two curves C_0 and C_0', then one curve will be a uniform magnification of the other.[4]

In (2.1), V is the scalar electric potential, V_0 the equivalent initial voltage of the particle, $\mathbf{A}$ is the magnetic vector potential, $d\mathbf{s}$ is an infinitesimal vector element of path length along the particle's trajectory, η is the charge to mass ratio of the particle, and k is a constant. The assumptions made in the introduction allow one to write $\mathbf{E}=-\nabla V$, and $\mathbf{B}=\nabla\times\mathbf{A}$, where $\mathbf{E}$ is the electric field intensity, and $\mathbf{B}$ is the magnetic flux density of the applied magnetic field.

Given, therefore, an electron-optical system, a corresponding light optical medium can be postulated, in accord with (2.1), and conversely. The terminology of light optics and electron-optics will be used interchangeably hereafter.

A region in space where V is constant is called homogeneous, if $\mathbf{A}=0$, isotropic.†

As was mentioned previously, an arbitrary set of points in object space, will be transformed, by the orbits of incident beam particles into a set of points in image space. It is necessary, for ideal focusing, that[5]

$$\int_{C_1} n_1 ds_1 = \int_{C_2} n_2 ds_2, \tag{2.2}$$

for every arbitrary object curve C_1, and its image curve C_2. The integrals in Eq. (2.2) are line integrals of the refractive indices n_1 and n_2 taken along the respective curves C_1 and C_2.‡

Without going into the details of the proof, it may be observed that Eq. (2.2) is a consequence of the requirements that the light rays shall satisfy Fermat's principle and that the mapping of object into image space shall be one to one and continuous. These requirements are just the ones to be satisfied by an ideal focusing system.

[4] H. Busch and E. Bruche, *Beitrage zur Electronenoptik* (Johann Ambrosius Barth, Leipzig, 1937), p. 29.

† A definition for isotropy is somewhat arbitrary, since the physically significant quantity is $\mathbf{B}$ rather than $\mathbf{A}$. When $\mathbf{B}=0$, then $\mathbf{A}=\nabla\phi$, where ϕ is an arbitrary scalar function. But if $\mathbf{A}$ is chosen to be zero when $\mathbf{B}=0$ then n' becomes independent of trajectory direction; this is consistent with the terminology of light optics.

[5] C. Caratheodory, *Geometrische Optik* (Julius Springer, Berlin, 1937), pp. 70–74, gives a general proof and a bibliography for this theorem. It is a version of Eq. (2.2) which was rediscovered by Gabor.

‡ Condition (2.2) cannot be applied to the paraxial equation. See Caratheodory, reference 5, for a discussion.

The discussion in the remainder of this note will be confined to an important special case, namely that of isotropic object and image spaces. Caratheodory§ has shown that under this condition Eq. (2.2) will be satisfied if and only if the transformation of object into image space is conformal.

Now all three-dimensional conformal transformations are given by[6]

$$x_2=\pm\frac{h(x_1-a)}{r^2},$$
$$y_2=\pm\frac{h(y_1-b)}{r^2}, \tag{2.3}$$
$$z_2=\pm\frac{h(z_1-c)}{r^2},$$

where x_1, y_1, z_1 are the rectangular coordinates of a point (in object space), x_2, y_2, z_2, the rectangular coordinates of the transformed point (in image space), h, a, b, c are constants, and $r^2=(x_1-a)^2+(y_1-b)^2+(z_1-c)^2$. The transformed differential line element, ds_2, is given in terms of the original one, ds_1, by

$$(ds_2)^2=\frac{h^2}{r^4}(ds_1)^2. \tag{2.4}$$

Equations (2.1)–(2.4) lead to the functional equation,

$$\frac{V_1(x_1, y_1, z_1)}{V_2(x_2, y_2, z_2)}=\frac{h^2}{r^4}. \tag{2.5}$$

Equation (2.5) is a necessary relationship between the potential functions in object and in image space, for ideal focusing. It cannot be used to deduce the potential distribution within the intervening lens.

Equation (2.5) can be used to determine the necessity for space charge in image, or in object space, in certain cases. Thus, if image space is homogeneous, then on differentiating Eq. (2.5) twice with respect to x_1, y_1, and z_1, and adding the derivatives, one finds that

$$\nabla_1^2V_1\neq 0, \tag{2.6}$$

where the subscript 1 on the Laplacian indicates differentiation with respect to the object space coordinates. Similarly, if the roles of V_1 and V_2 are interchanged with $V_1(x_1, y_1, z_1)=\text{const}$, then

$$\nabla_2^2V_2\neq 0. \tag{2.7}$$

It follows that space charge must be present, for ideal focusing, regardless of the properties of the intervening lens, if both object and image space are isotropic, and just one of the two spaces is homogeneous.

§ Reference 5, p. 75.

[6] G. Monge *Application de l'Analyse a la Geometrie* (corrigee et annotee par Liouville) (Bachelier, Paris, 1850), fifth edition, Note 6°, p. 609.

Equation (2.5) may be satisfied by systems with inhomogeneous object and image spaces which are free of space charge, as the example $V_1=-x_1/(x_1^2+y_1^2+z_1^2)^{+\frac{1}{2}}$, $V_2=-x_2/(x_2^2+y_2^2+z_2^2)^{+\frac{1}{2}}$ shows, since V_1/V_2 satisfies Eq. (2.5) and both V_1 and V_2 satisfy Laplace's equation, in object and image space, respectively.

Ideal focusing systems with inhomogeneous image and object spaces are discussed in the next two sections. As can be seen from the above arguments, some properties of the system as a whole must be known then before the need for space charge can be determined.

III. SPHERICALLY SYMMETRIC LENSES

Isotropic spherically symmetric lenses which ideally focus a three-dimensional object space onto a three-dimensional image space are considered in this section. The lenses studied are those which include object and image space, and which may be described by a single scalar potential, $V(r)$. It is shown that all such lenses must contain space charge, if they are to satisfy Eq. (2.5). Some spherically symmetric lenses are exhibited.

The general solution of Laplace's equation which is spherically symmetric is $V(r)=K_1/r+K_2$, where K_1 and K_2 are constants. The particle trajectories in the Coulomb field are conic sections.[7] The equation for any orbit may be written as

$$u=\gamma+(u_0-\gamma)\cos\theta+u_0'\sin\theta, \tag{3.1}$$

with $u=u_0$, $du/d\theta=u_0'$ at $\theta=0$, $u=1/\rho$, and ρ, θ plane polar coordinates. $\gamma=\gamma_1+\gamma_2u_0'^2$ where γ_1 and γ_2 are functions of u_0, K_1 and K_2, but not of u_0'.

Let a bundle of orbits intersect at the object point $(u_0, 0)$. The object point will have a conjugate image point (u_1, θ_1) whenever u_1 and θ_1 are independent of u_0' over a range of values of u_0'. If one differentiates Eq. (3.1) with respect to u_0', and uses the conditions that $\partial u_1/\partial u_0'=\partial\theta_1/\partial u_0'=0$, one obtains

$$2\gamma_2u_0'=-\cot\frac{\theta_1}{2}. \tag{3.2}$$

Differentiating Eq. (3.2) again, with respect to u_0', yields

$$4\gamma_2=\frac{\partial\theta_1}{\partial u_0'}\csc^2\frac{\theta_1}{2}. \tag{3.3}$$

Now γ_2 must be zero, unless $\theta_1=2m\pi$ (m an integer), the only exceptional possibility. Evaluating the indeterminate form, one obtains

$$4\gamma_2=\frac{\partial^2\theta_1}{\partial u_0'^2}\Big/\cos\theta_1=0, \tag{3.4}$$

when $\theta_1=2m\pi$. Thus γ_2 is always zero. But this implies that if ideal focusing is to exist, it may occur only if object space is the surface of a particular sphere, namely the one for which $\gamma_2=0$ and not a three-dimen-

[7] E. T. Whittaker *Analytical Dynamics* (Dover Publications, New York, 1944), p. 86.

sional region. Hence ideally focused three-dimensional spherically symmetric regions require the presence of space charge.

Necessary conditions for ideal focusing have been given, both in Sec. II, and in this section, but no systems have yet been discussed which are ideally focused. Theoretical work in geometrical optics has shown that some plane lenses are ideally focused; expressions for these plane lenses will be used now to generate formulas for spherically symmetric ideal focusing lenses.

Let $n(x, y)$ be the index of refraction of any plane ideally focused lens. Consider those lenses for which $n(x, y)=n(\rho)$. Such lenses may be thought of as being diametral plane representations of spherically symmetric lenses, and therefore, on substituting r for ρ, in $n(\rho)$, one obtains a spherically symmetric ideally focused lens.

Luneberg[8] has shown that the plane lenses,

$$n(x, y)=\frac{2|f'(\xi)|}{1+|f(\xi)|^2}, \tag{3.5}$$

are ideally focused, where $f(\xi)$ is an arbitrary analytic function of the complex variable $\xi=x+iy$. If $f(\xi)=\xi^m$, m an integer, then $n(x, y)=n(\rho)=2m\rho^{m-1}/1+\rho^{2m}$, and $n(r)=2mr^{m-1}/1+r^{2m}$. The best known lens of this class is the "Maxwell Fisheye"[11] with $m=1$.

IV. SOME CYLINDRICAL LENSES

It is the object of this section to derive a class of cylindrical lenses, which are free of space charge, from some ideally focused spherically symmetric lenses. The midplanes of the cylindrical lenses will be ideally focused.

Glaser[3] has derived a space charge free cylindrical lens from the Maxwell Fisheye; the following treatment is somewhat more general than his method.

Let the even function $V(\rho)$ be a known scalar potential function in a diametral plane of an ideally focused spherically symmetric lens. The diametral plane is considered to be the plane $z=0$ in an axially symmetric system, with the axial coordinate denoted by z, and the origin in the plane coincident with the point of intersection of the plane and the axis.

The potential $V(\rho, z)$ at any point in an axially symmetric system can be given, in terms of the potential on the axis, by[9]

$$V(\rho, z)=\frac{1}{\pi}\int_0^{\pi} g(z+i\rho\cos\theta)d\theta, \tag{4.1}$$

where $g(z)$ is the potential on the axis. The function $g(z+i\rho\cos\theta)$ can be expressed in terms of a double power series in ρ and z, and integrated term by term. One obtains

$$V(\rho, z)=\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^r g^{(2r+s)}(0)z^s}{s!}\left[\left(\frac{\rho}{2}\right)^r\cdot\frac{1}{r!}\right]^2. \tag{4.2}$$

Now $V(\rho, 0)=V(\rho)$ can be given by

$$V(\rho, 0)=\sum_{m=0}^{\infty}\frac{V^{(2m)}(0)\rho^{2m}}{(2m)!}, \tag{4.3}$$

and, therefore, by setting $z=0$ in Eq. (4.2) and equating coefficients of like powers of ρ, in Eq. (4.3) and the resulting Eq. (4.2), one finds that

$$g^{(2t)}(0)=\frac{(-1)^t V^{(2t)}(0)\,(t!)^2 2^{2t}}{(2t)!}, \tag{4.4}$$

for all positive integral t.

Equations (4.2) and (4.4) still do not determine $V(\rho, z)$ completely, since the $g^{(2r+s)}(0)$, with s odd, are unknown. One may choose $g(z)$ to be an even function, however. Equation (4.2) becomes then

$$V(\rho, z)=\sum_{r=0}^{\infty}\sum_{t=0}^{\infty}\frac{(-1)^r g^{(2r+2t)}(0)z^{2t}}{(2t)!}\left[\left(\frac{\rho}{2}\right)^r\cdot\frac{1}{r!}\right]^2, \tag{4.5}$$

where the coefficients $g^{(2r+2t)}(0)$ are known, from Eq. (4.4).

It follows, from the evenness of $g(z)$, that $E_z(\rho, 0)=0$. Hence if the axial velocity of a charged particle is zero, at some time, in the plane $z=0$, the particle will remain in that plane thereafter.

The equation $V(\rho, z)=$ const determines an equipotential surface, and if two such surfaces are replaced by conducting metallic sheets, raised to the potentials $V(\rho, z)=K_1$, and $V(\rho, z)=K_2$, then the annular strip $\rho_1\leq\rho\leq\rho_2$, in the plane $z=0$, will have the property of ideal focusing, where ρ_1 and ρ_2 are defined by $V(\rho_1, 0)=K_1$, $V(\rho_2, 0)=K_2$. Furthermore if these two surfaces intersect the z axis [i.e., if $V(0, z)=K_1$ and $V(0, z)=K_2$ have at least one real solution], then the potential distribution at $z=0$ will not be subject to end effect corrections if the metallic sheets are continued to the points of intersection with the axis.

The author wishes to thank a number of friends, and in particular Dr. A. T. Nordsieck, for comments on the manuscript.

[8] R. K. Luneberg, *Mathematical Theory of Optics* (Brown University, Providence, Rhode Island, 1944), p. 204.

[11] Reference 5, p. 75, and reference 8, p. 97, contain discussions of the Maxwell Fisheye.

[9] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge, 1935), fourth edition, Ex. 1, p. 399.

An Investigation of Boron Carbide

FRANK W. GLASER AND DAVID MOSKOWITZ, *American Electro Metal Corporation, Yonkers, New York*

AND

BENJAMIN POST, *Polytechnic Institute of Brooklyn, Brooklyn, New York*

(Received January 5, 1953)

Mixtures of boron and carbon were heated at approximately 2000°C; the reaction products have been studied by x-ray diffraction and by density and electrical measurements. The existance of a wide range of homogeneity of "boron carbide" has been established. No evidence of compound formation corresponding to B_4C could be detected. Similarities between the structures of boron and "boron carbides" indicate that the latter may best be described as solutions of varying amounts of carbon in a slightly distorted boron lattice.

INTRODUCTION

IN the course of an investigation of metal-carbon-boron systems, Glaser[1] observed a surprising instability of boron carbide in the presence of a number of transition metals or their carbides. At relatively low temperatures (1000–1100°C), the transition metals of the 4th, 5th and 6th groups of the periodic table react vigorously with boron carbide powders, of approximate composition B_4C, to form the metal borides. Other reactions with B_4C, of a similar nature, have been reported.[2] In these reactions the boron-carbide phase behaves more like a reactive mixture of boron and carbon, than like the extremely inert compound it is reputed to be.[3]

Various carbides of boron, ranging in composition from BC to B_6C, have been reported.[4] In one of the earliest investigations in this field, Moissan[5] had reported a compound of composition B_6C. Many subsequent references to B_6C are based on his work. Thus, in his well known textbook of inorganic chemistry, Ephraim[6] refers only to B_6C in his discussion of carbides of boron.

In 1934 Ridgway[3] described a "new" carbide of boron. On the basis of chemical analysis and physical measurements, he concluded that the formula was B_4C, and that, in addition, "there is no evidence to indicate that the compound phase has a range of composition because in all samples the (x-ray diffraction) lines have the same positions within experimental error."[3] However, in their discussion of binary borides, Latimer and Hildebrand[7] refer to this compound as "the so-called boron-carbide, B_4C." More recently, Clark and Hoard[8] have discussed the possibility of ranges of composition from B_4C to $B_{4.67}C$.

It is evident that little agreement exists in the literature concerning the stoichiometry of carbides of boron. A primary objective of the present investigation has, therefore, involved efforts to prepare and characterize "boron carbides" of widely varying compositions. It was hoped that as a result, some light would be shed on the observed reactivity of boron carbide.

EXPERIMENTAL PROCEDURE

Boron,* ranging in purity from 95 to 96 percent, was mixed in the desired proportions with Dixon Grade 8485 graphite powder. The mixtures were poured into graphite dies heated by direct conduction, to temperatures in the neighborhood of 2000°C. A Leeds and Northrup Speedomax recorder and Rayotube were used to control and record reaction temperatures.

Changes in composition, attributable mostly to oxidation of boron, with subsequent volatilization of B_2O_3 at temperatures above 1300°C, and also to some carburization of the mixed powders, occurred during the hot-pressing process. All samples were therefore analyzed chemically for their boron content, using Blumenthal's procedure,[9] and also for their carbon content, subsequent to the preparative procedures. The precision of the boron analysis by this method, as determined by repeated trials, is better than 0.5 percent. Carbon analyses were made using a Lindberg carbon determinator.

Specimens of the boron carbide phase were crushed to better than through 100 mesh for pycnometric density measurements. Resistivity data were computed from potentiometric measurements at room temperature of the voltage drop across 1.5 cm long sections of hot-pressed samples of uniform cross section.

[1] F. W. Glaser, J. Metals 4, 391 (1952).
[2] Nelson, Wilmore, and Womeldorph, "Refractory Bodies Composed of Boron and Titanium Carbides Bonded with Metals." Presented at 99th Meeting of the Electrochemical Society (1951).
[3] R. R. Ridgway, Trans. Am. Electrochem. Soc. 66, 117 (1934).
[4] N. V. Sidgwick, *Chemical Elements and their Compounds* (Oxford University Press, New York, 1950), p. 368.
[5] H. Moissan, Compt. rend. 118, 556 (1899).
[6] F. Ephraim, *Inorganic Chemistry* (Nordmann Publishing Company, New York, 1943), p. 181.
[7] W. M. Latimer and J. H. Hildebrand, *Reference Book of Inorganic Chemistry* (MacMillan Company, New York, 1940), p. 86.
[8] H. K. Clark and J. L. Hoard, J. Am. Chem. Soc. 65, 2115 (1943).
* This grade of boron was asserted by the supplier to be 99.3 percent pure. It was found on careful chemical analysis to contain only from 95–96 percent of boron, plus 0.6 percent carbon, 0.4 percent iron, and traces of other elements.
[9] H. Blumenthal, Anal. Chem. 23, 992 (1951).

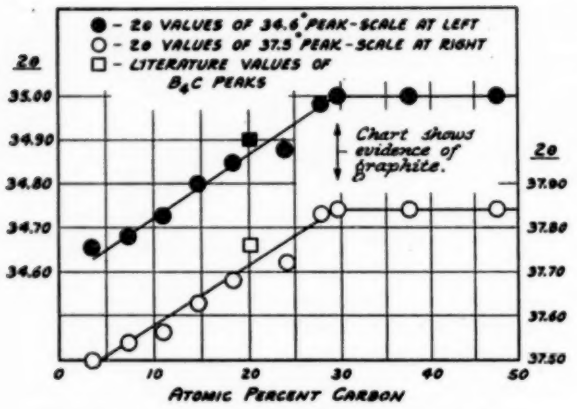


FIG. 1. 2θ values of the (10.4) and (20.1) reflections as a function of atomic percent carbon.

Copper K alpha-radiation was used in the x-ray diffraction investigations. X-ray powder film and spectrometer traces were obtained of all specimens, to identify the phases present. In addition, efforts were made to measure, with precision, certain 2θ values in the "boron-carbide" patterns. For this purpose, a Norelco wide range spectrometer, run at $\frac{1}{4}°$ per minute, was used. Preliminary x-ray studies showed marked broadening of lines, in most cases, at high Bragg angles. In addition, intensities of reflections fell off very rapidly for values of 2θ in excess of 50°. As a result, it was found that most accurate measurements could be made of the relatively sharp and intense (10·4) and (20·1) reflections, which occur at approximately 34.6° and 37.9° (2θ), respectively.

In these runs, portions of a batch of recrystallized sodium chloride were used as an internal standard. Strong sodium chloride reflections occur at 31.71° and 45.45° (2θ); these bracket the boron carbide lines under investigation. Measurements of the two boron carbide lines were reproducible to $\pm 0.02°$ (2θ).

RESULTS AND DISCUSSION

Analysis of the x-ray film and spectrogram patterns of the boron-carbon specimens indicated the following: The "normal" boron pattern[10] is observed in the case of specimens to which no carbon has been added. At a low carbon content (approximately 4 atomic percent), a change to the "boron-carbide" pattern occurs. Except for slight shifts in 2θ values of reflections, to be discussed below, this basic trend appears unchanged up to a carbon content of about 28 atomic percent. In this composition range no carbon lines could be detected. When the carbon content exceeds 28 percent strong reflections due to free graphite are observed.

The "boron-carbide" unit cell is rhombohedral;[8] the indexes used in this paper refer to the corresponding hexagonal unit cell. In Fig. 1, values of 2θ of the (10·4) and (20·1) reflections are plotted as ordinates against atomic percentages of carbon, as determined by final analysis.

It is evident from Fig. 1 that the variation of the spacings of the (10·4) and (20·1) reflections with carbon content is regular and almost linear. There is no indication of compound formation at 20 atomic percent carbon.

The calculated decrease in volume of the unit cell in going from 4 to 28 atomic percent carbon is 2.2 percent.

Efforts were also made to determine whether compound formation occurred at a boron to carbon atomic ratio of 4 to 1 by means of electrical resistivity and density measurements. Experimental difficulties were considerable. Since single crystals of a size suitable for resistivity measurements could not be obtained, hot-pressed specimens of uniform density were used; these were approximately $\frac{1}{4}$ in.$\times\frac{1}{2}$ in.$\times 1$ in. in size. Several measurements were made, using different samples for each composition, and the averages of these measurements are plotted in Fig. 2.

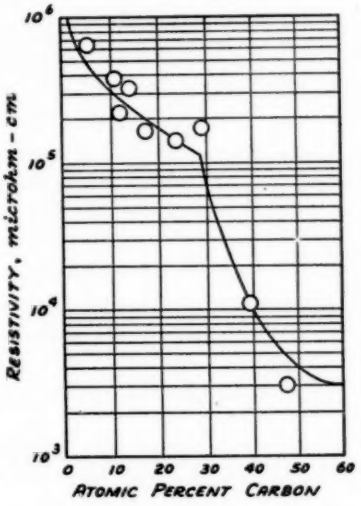


FIG. 2. Electrical resistivity of boron-carbon samples as a function of percent of added carbon.

There is no indication in Fig. 2 of any discontinuity or significant change of slope at a carbon content of 20 atomic percent. An abrupt decrease in resistivity is observed in specimens containing free carbon.

The results of pycnometer density measurements of a number of samples of varying carbon content are shown in Fig. 3. Here, too, no evidence of compound formation at 20 atomic percent can be detected. Specimens increased in density steadily from 2.29 g/cc, at zero carbon content, to 2.63 g/cc at about 24 atomic percentage of carbon.

According to Zhdanov and Sevastyanov[11] and Clark and Hoard,[8] the boron carbide structure can be represented formally as a distorted NaCl type of structure. Compact groups of twelve boron atoms situated at the vertices of nearly regular icosahedra, substitute for the sodium atoms, and linear groups of three carbon atoms

[10] A. W. Laubengayer *et al.*, J. Am. Chem. Soc. **65**, 1924 (1943).

[11] G. S. Zhdanov and N. G. Sevastyanov, Compt. rend. acad. sci. U.R.S.S. **32**, 432 (1941).

substitute for chlorine atoms. Clark and Hoard[8] have pointed out that additional "holes" between boron icosahedra are present in the boron carbide structure. These, they state, provide enough space for two additional boron atoms. Filling of these holes with boron atoms would tend to a B/C ratio of 4.67 to 1, thus accounting for reported instances where the B/C ratio exceeded 4.

It is interesting to note that filling of these "holes" with carbon atoms would lead to a B/C ratio of 12 to 5; this would represent the upper limit of carbon content. Results of the present investigation indicate that the "boron-carbide" structure can accommodate a maximum of about 28 atomic percent of carbon, in very good agreement with this 12 to 5 ratio. When the carbon content exceeds this value, the presence of free graphite is clearly indicated by x-ray, resistivity, and density measurements.

Hoard and Geller[12] have recently completed a determination of the crystal structure of boron. They report that the boron atoms are arranged in groups of twelve at the vertices of nearly regular icosahedra. They describe "the strong network structure (in both B and B_4C) in which surprisingly large holes are interspersed among the very compact icosahedral groups."

It appears, as a result of the present study, that the introduction of even small amounts of carbon, perhaps even less than 4 atomic percent, results in a rearrangement of the boron icosahedra from the "boron" to the "boron carbide" arrangement. The latter structure resembles more a solution of carbon among the boron icosahedra, with an upper solution limit of about 28 atomic percent of carbon, than it does a "compound" in the usual sense.

The density measurements are of some interest in this connection. It is clear that a range of composition from, say $B_{20}C$ to $B_{12}C_5$, accompanied by a 2.2 percent decrease in volume, would lead to an increase in density of approximately 36 percent, compared with the 15 percent increase which is observed (2.29 to 2.63). It is evident that, at low carbon concentrations, the icosahedral holes are occupied by a mixture of boron and carbon atoms. The increase in carbon content, therefore, appears to represent, in part at least, a replacement of boron atoms in icosahedral holes, by carbon atoms. The observed decrease in volume which accompanies the increase in carbon content may, therefore, reflect a replacement of boron atoms by smaller carbon atoms. The volume decrease does not necessarily imply bonding forces of considerable magnitude between the carbon and boron atoms. It is believed that this picture of the boron carbide structure as resembling a solution of carbon (and boron) in a boron lattice, helps explain the reactivity of "boron carbide" which was mentioned at the beginning of this paper.

[12] J. L. Hoard and S. Geller, J. Am. Chem. Soc. **73**, 1892 (1951).

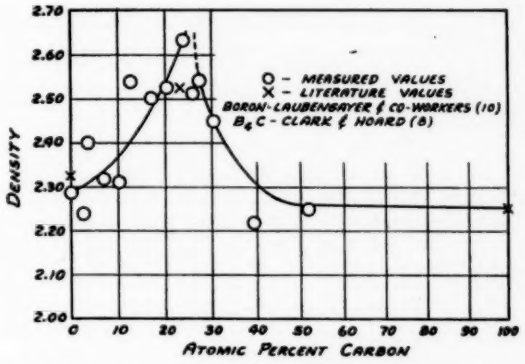


FIG. 3. Pycnometric density *vs* atomic percent carbon for boron-carbon compositions.

SUMMARY

It is believed that the observed reactivity of "boron carbide" with transition metals of the 4th, 5th, and 6th group of the periodic table, as well as with Fe, Ni, Co, and Th, can be attributed, at least in part, to the nature of the boron carbide structure which resembles a solution of carbon (and boron) in a boron lattice. This solution of carbon in boron has a range of at least 4 to 28 atomic percent and is accompanied by a 2.2 percent volume decrease over this range (with increasing carbon content to the maximum of about 28 percent). Density measurements appear to indicate that an increase in carbon content leads, in part at least, to a replacement of boron atoms by carbon atoms in icosahedral holes.

Lattice parameter, electrical resistivity, and density measurements have not produced any evidence to support the existence of the previously reported, stoichiometric compound B_4C at 20 atomic percent carbon. A rather wide range of homogeneity for this type of boron carbide was indicated by the aforementioned test methods, which appeared to reveal the existence of free graphite only for compositions containing more than 28 to 30 atomic percent of carbon added to boron.

ACKNOWLEDGMENT

The authors wish to thank the U. S. Office of Naval Research, under whose sponsorship this research was conducted, for permission to publish these data.

They also gratefully acknowledge the valuable assistance of W. Ivanick and V. Gizzi during sample preparation and physical measurements.

The Creep of Zinc Single Crystals under Direct Shear*

ELLIOT H. WEINBERG†
Department of Physics, State University of Iowa, Iowa City, Iowa
(Received January 16, 1953)

Room temperature measurements on very pure specimens verify the Tyndall law, $s=at^m$, which was determined by tensile measurements. The value of m is approximately 0.6. Discrepancies in tensile creep tests reported by other observers are shown experimentally to be caused most probably by bending of the specimen during the creep process. The Bauschinger effect has also been investigated and creep at the very small stresses needed appears to follow the same creep law.

INTRODUCTION

A WEALTH of technical information has been accumulated in the past concerning the creep phenomenon in polycrystalline materials of engineering interest and in various noncrystalline organic materials, all at large strain values.[1] Sully[2] states that "The first systematic investigation of the creep of a variety of metals and alloys over a range of stresses and temperatures was made by Andrade." Andrade[3] actually measured the creep as a function of time under constant tensile stress for polycrystals of lead, copper, cadmium, tin, and mercury, as well as various brasses. For the pure metals he found a law of the form

$$L=L_0(1+\beta t^{\frac{1}{3}})\exp(kt), \tag{1}$$

in which the constants β and k are empirical, differing for different materials as well as for the same materials with different histories. For small strains E this may be rewritten as

$$E=\beta t^{\frac{1}{3}}+kt. \tag{2}$$

The first term is generally interpreted as the transient strain, as its rate of change decreases with time, whereas the second, with a constant strain velocity, is called the steady state term. In the literature[2] one finds creep as a function of time divided into regions, a primary one in which the strain velocity is initially high but decreases rapidly, a second region in which the velocity is constant, and sometimes a third region of rapidly increasing velocity. Andrade's equation accounts for the first two regions.

Although Andrade's law was obtained for polycrystals, support has been found for it for single crystals as well. Using single crystals of zinc in the form of wires two millimeters in diameter, Cottrell and Aytekin[4] find that their data agree with a law of the form

$$L=L_0(1+\rho t^{\frac{1}{3}}+kt). \tag{3}$$

Their strain region started at about five percent and extended to forty percent. The three empirical constants are not the same for different crystals of zinc.

Another article[5] on the creep of single crystals, this time for aluminum, offers this relationship

$$L=L_0-1/(bt+c)+kt. \tag{4}$$

Here three empirical constants are introduced and applied for one crystal for five different stresses.

Measurements on single crystals of zinc in the very low strain region, where presumably little chance exists for deformations such as bending and twinning to take place, have been made in this laboratory by Tyndall,[6,7] who has proposed a law of the form

$$S=at^m \tag{5}$$

in which S is the strain under a given stress, t is the time, and a and m are empirical constants. The constant m is approximately 0.5 at room temperature for all crystals, while a is a function of the stress, but may differ from crystal to crystal and is history-dependent for any one crystal.

Slifkin and Kauzmann[8] have made measurements on creep in the low strain region also and find that their results agree with the equation,

$$\text{strain rate}=at^{-n}, \tag{6}$$

with a and n empirical constants. This expresses the same facts as Eq. (5) with $n=1-m$ and a taking the place of the product am, since differentiation of Eq. (5) gives $dS/dt=amt^{m-1}$. Although the values of n found by Slifkin and Kauzmann vary considerably from $+0.46$ to -0.28, many are in the neighborhood of 0.7, which corresponds to 0.3 in Eq. (5).

It is, perhaps interesting to note that a law of this type was proposed by Nutting[9] as a result of his

* A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics in the Graduate College of the State University of Iowa.

† Now at University of Minnesota, Duluth Branch, Duluth, Minnesota.

[1] C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1942).

[2] A. H. Sully, *Metallic Creep* (Butterworth Scientific Publications, London, 1949).

[3] E. N. DaC. Andrade, Proc. Roy. Soc. (London) **A84**, 1 (1910).

[4] A. H. Cottrell and V. Aytekin, Nature **160**, 328 (1947).

[5] R. R. Hasiguti, Nature **168**, 706 (1951).

[6] E. P. T. Tyndall, "Symposium on Plastic Deformation of Crystalline Solids," Pittsburgh, May, 1950, p. 49 (issued by U. S. Office of Naval Research).

[7] E. P. T. Tyndall, J. Appl. Phys. **21**, 939 (1950).

[8] L. Slifkin and W. Kauzmann, J. Appl. Phys. **23**, 746 (1952).

[9] P. G. Nutting, Proc. Am. Soc. Testing Materials **21**, 1162 (1921).

observations on the creep of amorphous organic materials. His equation was

$$S=bt^m\sigma^n, \tag{7}$$

in which S is the shear strain, σ the stress and b, m, n empirical constants differing for different materials.

In all of the work on creep as a function of time and, with one exception,[10] on the stress strain relation for single and polycrystals, the method of measurement has been tensile, that is, the crystal has been supported at one end of its length and the load has been applied at the other. In the case of zinc, it has been found that if one resolves the stress thus applied into a shear along the basal plane (0001) in one of the $[11\bar{2}0]$ directions, and a normal stress, then the behavior of the crystal is affected only by the magnitude of the resolved shear stress for normal stresses varying by a factor of 2500.[11]

In order to test the effectiveness of shear along the basal plane in producing creep, without the large accompanying normal stresses which are present in tensile measurements, an apparatus was devised to measure the creep of a single crystal of zinc under direct shear, so that the stresses were parallel to the basal planes and to the length of the specimen.

Although successive compressive and tensile tests have on a few occasions[12] been carried out on single crystals, the difficulties of observing creep in the reverse direction are manifest in any tensile measurements. Since considerable interest is attached to the effect of strain in one direction on the creep process in the reverse direction, and since with a direct shear apparatus it is extremely simple to reverse the shear direction, attention was devoted to this problem also.

Since this apparatus permits a readily measurable direct shear to be applied on any plane desired (by using a crystal grown with an appropriate orientation) and at the same time restrains creep along other planes and directions, it was feasible to investigate also the possibility of creep along other crystallographic planes and directions, as well as the possibility of twinning[1] a single crystal, since, for zinc, the twinning plane is well known.[13,14]

EXPERIMENTAL PROCEDURE

Crystals in the form of rods with a square cross section of about 50 square millimeters and a length of 15 centimeters were grown in air in a transite mold by a method used frequently in this laboratory.[15] The materials used were Bunker Hill (BH4),[16] a commercial grade of zinc with less than 0.01 percent impurity, New Jersey Zinc Company Spectrographically Pure (SP) zinc, with less than 0.001 percent impurity and a third lot designated R212.[17]

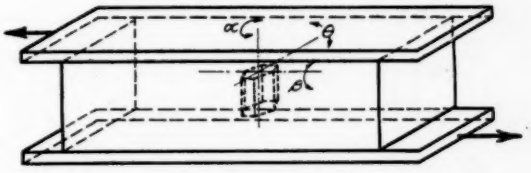


FIG. 1. Schematic showing a θ of 90°, α of 0°, β of 0° crystal under shear. The hexagon (in the basal plane) is parallel to one pair of faces ($\beta=0°$), its normal is perpendicular to the length of the specimen ($\theta=90°$), and one side of the hexagon is in the direction of the length ($\alpha=0°$). Changes in these angles are produced by rotation about each of the axes shown.

For the basic investigation, crystals were grown with a rectangular face parallel to the basal plane and with the $[11\bar{2}0]$ direction along the length of the rod. This orientation is completely described by three angles: $\theta=90°$, $\alpha=0°$, and $\beta=0°$. The angle θ is the angle between the normal to the basal plane and the length of the rod, α is the angle between the $[11\bar{2}0]$ direction in the basal plane and the rod length projected, if necessary, into that plane, and β is the angle between the intersection of the plane to be sheared (i.e., usually the basal plane) with the face of the crystal rod and the edge in that face; e.g., a β of 0° indicates that the shear plane is parallel to one pair of rectangular faces of the crystal (see Fig. 1).

For all crystals investigated, the procedure after growth was to saw the crystal to an appropriate length using a fine jeweler's saw, after waxing the crystal into place in a holder of channel iron. After sawing, the crystal was then etched in dilute HCl to remove the cold work caused by the sawing operation as well as to remove the oxide which would interfere with the mounting operation. Thereafter, the pair of faces parallel to the plane to be sheared were "soldered" with Wood's metal between two steel plates, and the whole could then be mounted between a pair of shearing plates, which were readily attachable by means of bolts.

The actual load was applied by using standard weights suspended from a wire which went over a pulley and pulled horizontally on the upper plate to which the frame was bolted, while the lower remained rigidly attached to an I beam. A spring scale was inserted into the wire between plate and weight so that any frictional binding in the pulley system could be detected.

Angular strain, or shear, was measured directly by the rotation of a mirror, with one support on each of the two crystal holding plates, or occasionally, for comparison and when circumstances permitted, with both supports directly on the end face of the crystal. The mirror was placed behind a telescope objective lens of about 100-cm focal length which focused the image of

[10] K. Bausch, Z. Physik **93**, 479 (1935).
[11] E. Schmid and W. Boas, *Plasticity of Crystals* (English translation) (Hughes and Company, Ltd., London, 1950).
[12] G. Sachs and H. Shoji, Z. Physik. **45**, 776 (1927).
[13] C. H. Mathewson and A. J. Phillips, Proc. Amer. Inst. Min. Met. Eng., Met. Div., 445 (1928).
[14] N. Thompson, and D. Millard, Phil. Mag. **43**, 423 (1952).
[15] C. A. Cinnamon, Phys. Rev. **46**, 215 (1934).
[16] This particular lot (BH4) consisted of a 50-pound slab kindly donated to the laboratory by Mr. Walter P. Carrol, Mgr, National Lead Company, Chicago, Ill.
[17] Kindly donated by the New Jersey Zinc Company and stated to be very nearly the same purity as the SP zinc.

UNIVERSITY OF MICHIGAN LIBRARIES

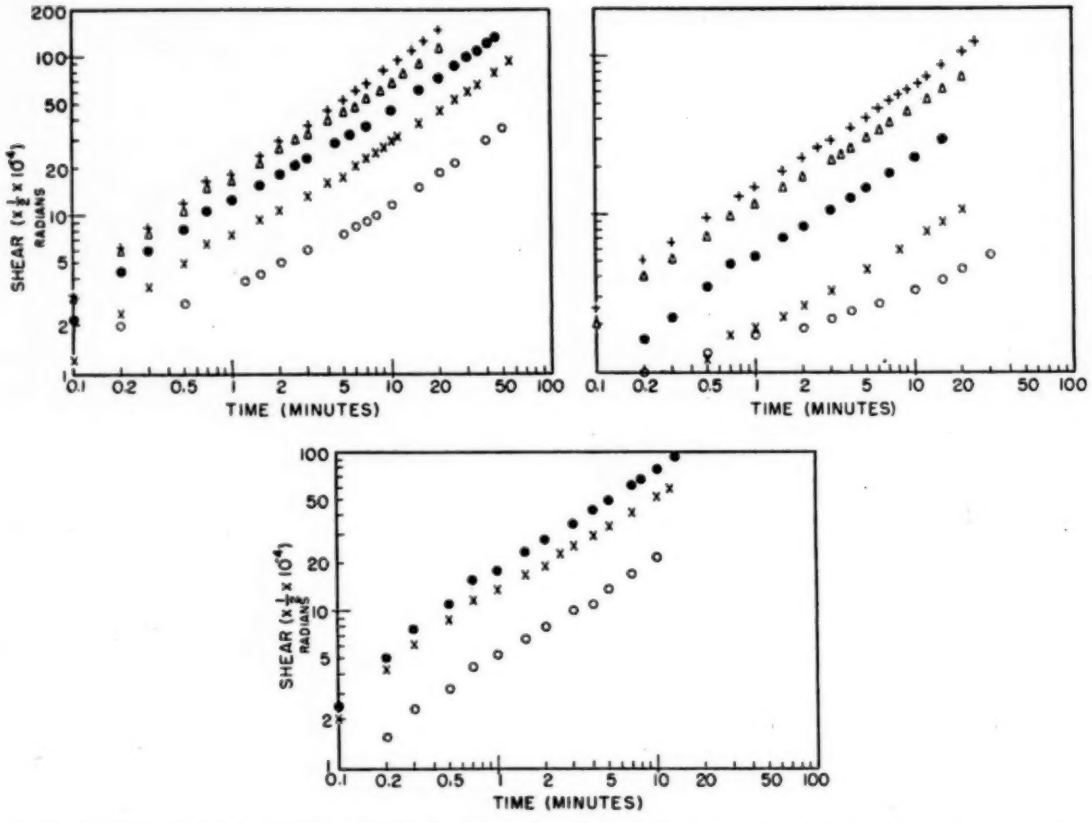


FIG. 2. Creep *vs* time for crystal 3A on three consecutive days, upper left, upper right, and lower, respectively.

Symbol	Top Left	Top Right	Bottom
○	37.2	35.0	39.4 g/mm²
×	39.4	37.2	41.6 g/mm²
●	41.6	39.4	43.8 g/mm²
△	43.8	41.6	
+	46.0	43.8	

a one-mil hairline onto a scale having 100 divisions per cm. The passage of the hairline image across the scale was then observed with a low power microscope[18] mounted on a calibrated traveling carriage, so that deflections larger than the range of divisions within the field of view of the microscope could also be measured.

One division on the scale corresponded to a shear of 50 microradians, and since tenths could generally be estimated, angles of 5 microradians were detectable. All parts of the system, i.e., the crystal supports, the hairline, lens, and microscope and scale, were fastened to the same I beam at different points, so that motion of the I beam as a whole would not interfere with the creep run.

Except where particularly noted, creep-time runs were made by a standard procedure as follows: As soon as a new load increment was added, the stop watch was started, and scale readings were taken through the microscope at definite time intervals. Thus the zero of time was that time when the load increment had been completely added. The scale zero was obtained, however, by linearly extrapolating the first two readings to zero time, thus removing to a large degree the effect

TABLE I. Composition and orientation of the crystals with their yield points. BH[a]=Bunker Hill Zinc Company 99.99+ percent zinc; SP and R212=New Jersey Zinc Company 99.999 percent zinc. Angles α, β, and θ are defined in Fig. 1 except for crystals 5, 6, 7, in which ρ is the angle between the normal to the twin plane and the length of the specimen.

Crystal	Zinc	θ	α	β	Yield Point (g/mm²)
1	BH	83°	10°	0°	82
2	BH	90°	10°	15°	96
3	BH	90°	0°	0°	A 28 B 22
4	BH	87°	17°	0°	29
5	BH	43°	30°	0°	>35 (Fracture)
6	BH	43°	30°	0°	105
7	BH	43°	30°	0°	100
8	R212	90°	30°	0°	24
9	SP	90°	0°	0°	12
10	BH	poly crystal	—	—	>100

[18] Sensitive optical lever, originally described by Roberts, *Heat and Thermodynamics* (Blackie, London, 1940).

of the slight scale movement which unavoidably must occur during any loading process which requires a finite loading time. All data were obtained at room temperature.

RESULTS

Yield Point

The general behavior of a single crystal specimen of zinc is to give small sudden (elastic) strain increments as the load is increased by steps from zero. As the stress needed to initiate creep is approached, these sudden strain increments become noticeably larger and no longer completely elastic for the same stress increments as have been previously applied. There is, at this time, no perceptible further strain thereafter, until, finally, a load is reached for which the sudden strain increment is now followed immediately by a continued strain increase at a measurable rate over an extended period of time. This minimum stress at which creep

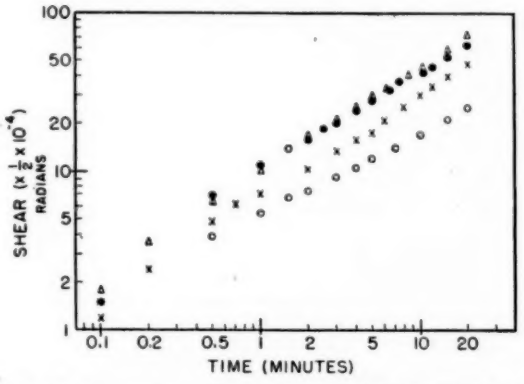


FIG. 3. The creep-time relation for crystal 3B. The 33.8-g/mm² run shows the effect of using too small an increment. ○ = 27.2 g/mm², × = 29.4 g/mm², ● = 31.6 g/mm², △ = 33.8 g/mm².

noticeably begins will be called the yield point. The value of this yield point depends on the apparatus employed, since, with a more sensitive device, it might be possible to detect time-dependent creep which would not be detectable with the present apparatus in any reasonable period of time.

Table I gives the yield point for virgin crystals as well as the orientation of each specimen and the type of zinc of which each specimen is composed. The values shown for crystals properly oriented for shear along their basal planes (i.e., crystals 3, 4, 8, 9) are in good agreement with other values reported,[7,12] after the tensile stresses have been resolved along the basal plane.

Form of the Creep Curve

As can be seen in Figs. 2–6, the plot of log strain against log time is always a straight line when the successive load curves are taken under the standard procedure, but where other attending circumstances are quite variable. Thus for all initially undeformed

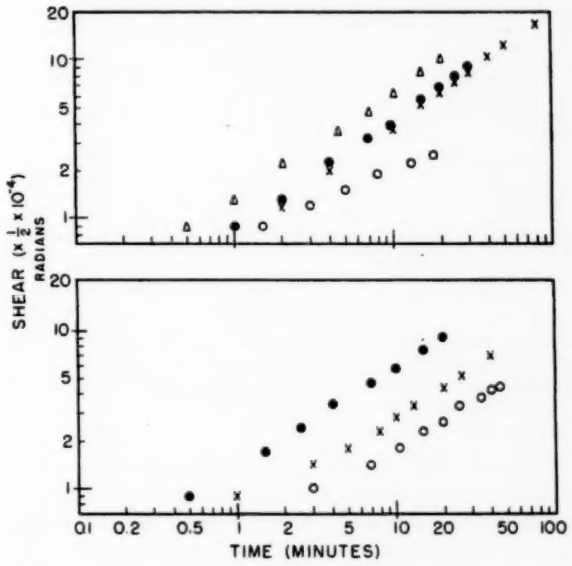


FIG. 4. Creep *vs* time for crystal No. 4. The lower runs were made by reversing the stress within 30 minutes after completing the runs shown in the upper box. The solid circles in the upper box also show the effect of adding too small an increment.

Symbol	Upper	Lower
○	34	36 g/mm²
×	45	49 g/mm²
●	48	60 g/mm²
△	56	

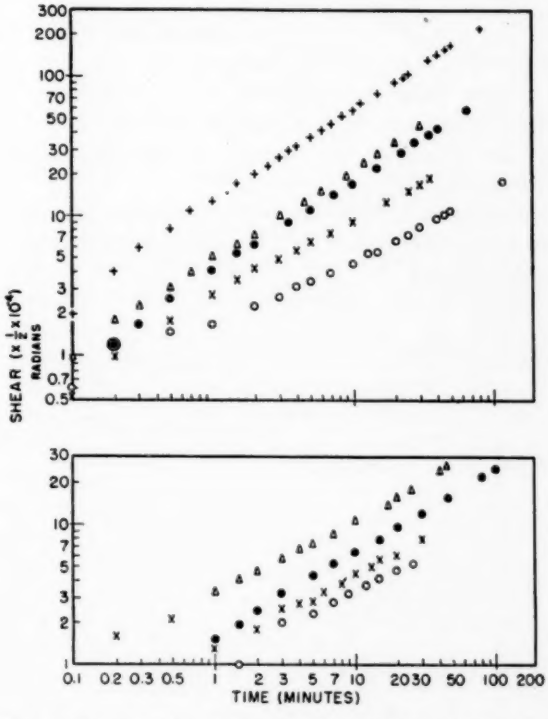


FIG. 5. Creep *vs* time for crystal No. 8A. The lower set were made by reversing the stress used in the upper set. The 20 g/mm² run was made 15 minutes after completing the upper set. The remainder of the lower runs were made 12 hours later.

Symbol	Upper	Lower
○	28	20 g/mm²
×	32	26 g/mm²
●	36	28 g/mm²
△	40	32 g/mm²
+	48	

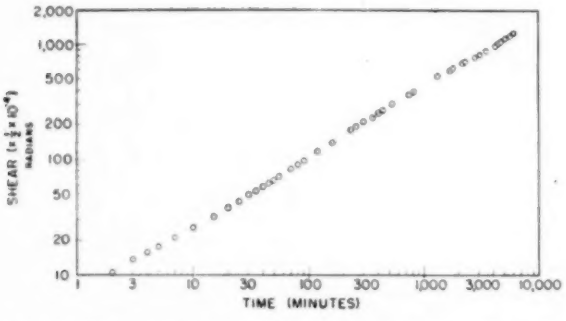


FIG. 6. The creep-time relation for crystal No. 9A under a stress of 24 g/mm². Note the regular continuation of the run into the macroscopic region. The strain scale here starts at 10 divisions instead of the usual one.

crystals of the proper orientation ($\alpha=0°$, $\beta=0°$, $\theta=90°$) the creep curve is a very good straight line with a slope between 0.5 and 0.6. For crystal 3A, the average slope of 15 runs is 0.54, with the extremes all within the range 0.5 to 0.6. On the first direct runs made on crystal 3B the slopes ranged from 0.5 to 0.7. The over-all average on this crystal, including those runs both in the forward and in the reverse direction is 0.5, with the lowest slopes of $m=0.3$ being obtained on the first runs made after stress reversals and the highest slopes, of about 0.7, being obtained on the last runs (i.e., those of highest stresses) in any sequence. Similarly, the slopes for crystal 4 centered about 0.62 with a range from 0.4 to 0.7. For crystal 8A, m varied from 0.5 to 0.65 with about half of the runs made having slopes approximately 0.59. For crystal 9A, for which only the long term run shown in Fig. 6 was made, the slope was 0.59.

For the same crystal run under a number of loads for three successive days, Fig. 2 shows the similarity of the results obtained.

Figure 3 shows a sequence of runs made on another crystal on one day. The effect of too small a load increment is shown by the last load increment applied.

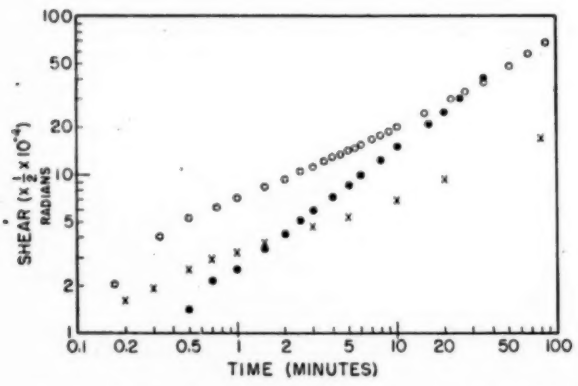


FIG. 7. The creep-time relation for crystals not correctly oriented for basal shear. Note the high stresses needed to initiate creep. Note also the pronounced hardening of crystal No. 2 in the single day elapsing between the 96 g/mm² and the 170 g/mm² runs. ○ = crystal No. 1, 102 g/mm², ● = crystal No. 2, 96 g/mm², × = crystal No. 2, 170 g/mm².

In Fig. 4 the lower set of curves was obtained by reversing the stress 30 minutes after completing the upper set of runs.

In Fig. 5 the first (bottom) run in the lower set was made 15 minutes after reversing the stress used in obtaining the last (top) run in the upper set. Here creep was obtainable at 20 g/mm², a load considerably lower than the minimum load needed to produce creep in the initial forward direction. The remaining runs in the lower set were taken about 12 hours later, and the loads necessary to produce creep at the same rate are now about the same as those in the upper set, even though the stress is still reversed.

Figure 6 shows the validity of the creep law for a run continued over a long period of time and to a shear of 7.5 percent.

In fact, for all single crystals observed, even those whose orientation was considerably off (see Figs. 7 and 8) from that for basal shear, thus requiring much larger stresses, no evidence was obtained in favor of the existence of two or three different successive creep processes. As Fig. 7 shows, the two runs on crystal number 2, taken a day apart, with the 96 g/mm² run on the first day, show a dissimilarity not found in the ideally oriented crystals. Not only are the slopes very different, but the load needed to get creep curves in approximately the same strain region is much greater on the second day. This may be interpreted as hardening.

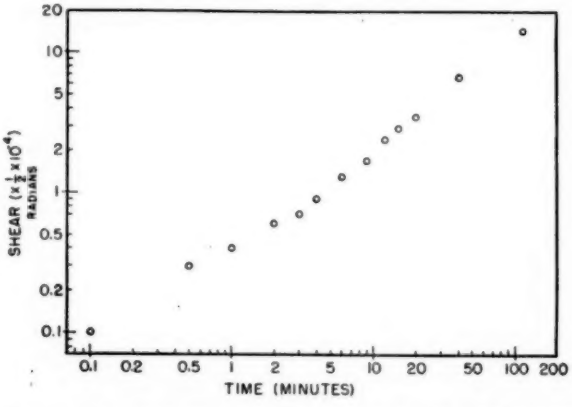


FIG. 8. The creep-time relation for crystal No. 3B mounted with its basal plane perpendicular to the shear direction (i.e., the crystal has been rotated through 90° about its length and remounted). The strain scale here starts at 0.1 division instead of the usual one. The slope of the line is about 0.8. Creep here must be interpreted as shear along the prismatic planes, since the basal planes are rigidly held.

In Fig. 8, which exhibits the creep curve for a previously strained crystal (3B), the same general law again appears to hold, but with m much different and a somewhat more irregular shape to the curve. Creep here must be interpreted as shear along the prismatic planes since the basal planes are rigidly held.

An interesting deviation from the regular procedure

described so far can be seen in Fig. 9. It was obtained by shearing the crystal at a particular load (open circles o), then unloading, and very rapidly reloading it again to the same load value. The curve obtained is shown by the × points in the figure. It not only lies below the previous run at the same load, but is far from being a straight line. The upper end is, however, approaching parallelism with the first run. These last data are shown in a different fashion by the solid circle points which are plotted with time and strain reckoned from the beginning of the open circle line. Thus it is seen that the removal of the load for a short time has checked the creep process so that it goes on at a generally lower rate than it would have had the load been left on continuously.

For a constant slope m on a log-log plot, the values of the coefficient a must be functions of the stress applied, and it might be hoped that some insight into

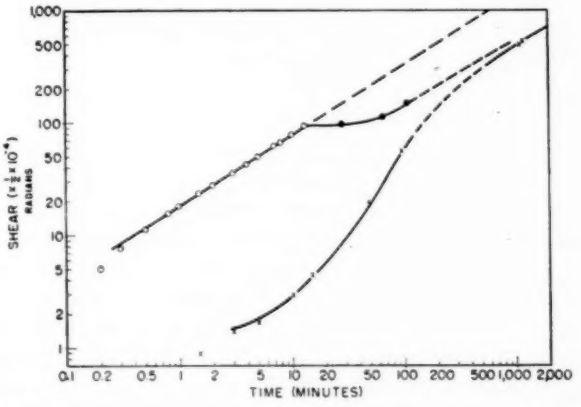


FIG. 9. The creep-time relation for crystal No. 3A at 43.8 g/mm². The lower curve was obtained by unloading and immediately reloading the crystal to the same value. The solid circles indicate how the lower set of points would appear if the time and strain were plotted as a continuation of the upper curve, instead of taking a new zero for strain and time.

the creep mechanism could be gotten in this way. Several attempts have been made to obtain this relation between stress and the a coefficient. Tyndall[6,7,19] finds that his data fit a law of the form

$$a=\exp(c\sigma).$$

As there is some theoretical evidence[20] that the relationship ought to fit an equation of the form

$$a=\exp(-k/\sigma),$$

Figs. 10, 11, and 12 show both types of plot. Evidently there is little reason to prefer one over the other, a consequence of the fact that the stress values cover too narrow a range to discriminate between the two. Actually a linear relationship between a and σ or $(\sigma-\sigma_0)$ as well as powers of either can be fitted as well to the data presented, with σ_0 as an adjustable constant.

[19] T. A. Read and E. P. T. Tyndall, J. Appl. Phys. **17**, 713 (1946).

[20] J. S. Koehler, Phys. Rev. **86**, 52 (1952).

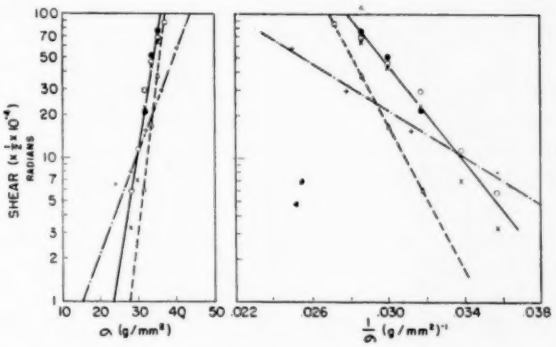


FIG. 10. Plot of log strain at the end of 10 minutes (a quantity proportional to the coefficient a) as a function of the applied stress σ for crystal No. 3A for several runs. In the left-hand plot the abscissas are the stresses in g/mm², while on the right they are the reciprocals of these values. Left plot: ○ = initial run, × = second day, ● = third day, △ = after a 5-day lapse, total shear about 0.1 radian, + = reverse 2 days later.

In addition to the small stress range limitation, the situation is complicated by the effect of putting on too small an increment. If, consistently, a large enough increment is used (where large enough is defined by the resulting effect) each run appears parallel and higher on the log S, log t plot. If, however, the increment is too small, as noted previously in reference to Figs. 3 and 4, it causes the coefficient a to be about the same or even lower than the previous one though this effect is usually removed by adding a sufficiently higher load on the next following run. Thus any regular increase of a with increasing stress depends on the procedure used in getting the data.

The history dependence of the differences in behavior of the a *vs* σ plots makes the finding empirically of any definite one to one relationship meaningless. For example, in Fig. 11 the open circles near the top appear to flatten out the trend of the lower points. This "increment effect" is not readily observable, however, in Figs. 10 or 12. The over-all regularity of these curves should be noted, nevertheless, since

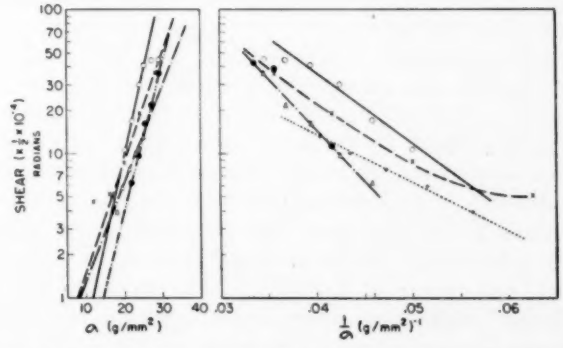


FIG. 11. Plot of log strain at the end of 10 minutes (a quantity proportional to the coefficient a) as a function of the applied stress σ for crystal No. 3B for several runs. In the left-hand plot the abscissae are the stresses in g/mm², while on the right they are the reciprocals of these values. Left plot: ○ = initial run, × = immediate reverse, ● = reverse continued after 2 days, △ = forward after 1 day.

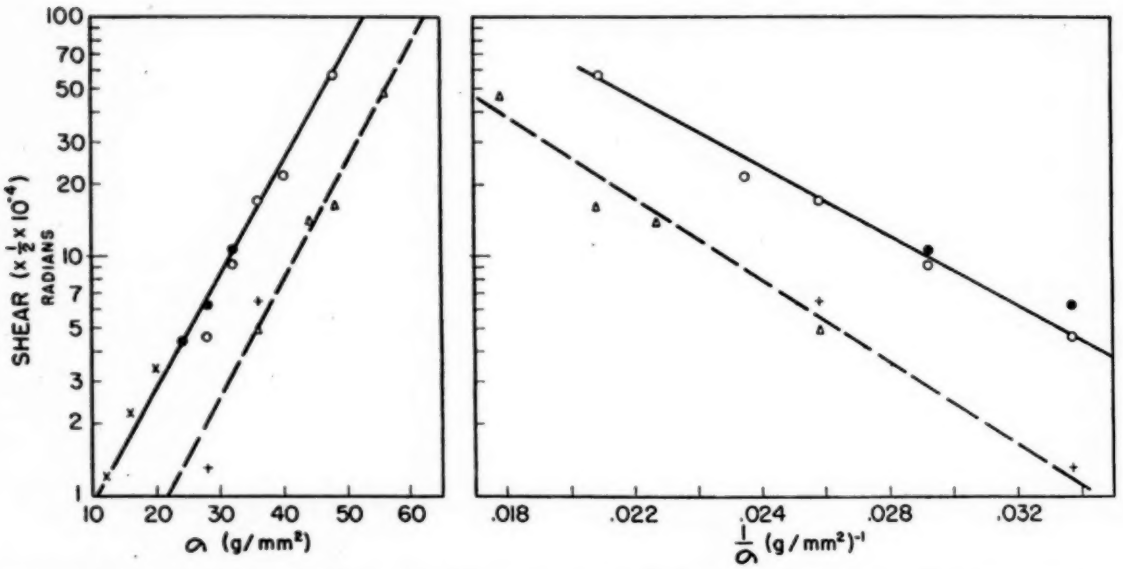


FIG. 12. Plot of log strain at the end of 10 minutes (a quantity proportional to the coefficient a) as a function of the applied stress σ for crystal No. 8A for several runs. In the left-hand plot the abscissae are the stresses in g/mm², while on the right they are the reciprocals of these values. Left plot: ○ = initial run, × = immediate reverse, ● = reverse continued after 12 hours, △ = immediate forward, + = forward continued after 24 hours.

over the stress ranges here employed, the regularity is much more significant than the history-dependent deviations. In fact, as will be further discussed in the next section, it is noteworthy that not only runs in one direction of shear give qualitatively the same a *vs* σ curves (despite history-dependent variations), but very similar data are obtained even with the stress reversed after sufficient anneal time.

Creep Recovery and the Bauschinger Effect

The Bauschinger effect[21] is the observed lowering of the yield point, or elastic limit, after previous plastic strain in the reverse direction. It is the opposite of work hardening which is the raising of the yield point after previous strain in the same direction. For polycrystalline materials the needed reverse stress may even become negative, in the sense that a considerable amount of plastic recovery is supposed to take place over an extended period of time, i.e., the specimen creeps under zero external load in the opposite direction after the original load is removed. This effect is extremely minute in single crystals, as it should be, since local stress inhomogeneities are much less likely to develop in a single crystal. For this reason it was believed for a long time that the Bauschinger effect was observable only in polycrystals, but there is, at least, one report of this effect in a single crystal of α-brass.[12] With this exception, it appears to be the general opinion that single crystals do not exhibit this Bauschinger effect.

On the other hand, the observation of well-marked hysteresis in a crystal following a permanent strain[19] is undoubtedly related to the Bauschinger effect in that, particularly in the unloading branch of a hysteresis loop, the strain decreases in greater proportion to decrease of load than that shown by a previous "elastic" curve. That the Bauschinger effect has not been more often directly observed is probably a result of the fact that, as previously mentioned, the majority of plastic studies have been carried out with tensile apparatus so that alternate stress reversals were extremely difficult, particularly when thin wire crystals were used.

Table II exhibits the Bauschinger effect as well as the opposite phenomenon of strain hardening. The first column is merely the identifying number of the crystal. In the second column is shown the direction of the stress applied, i.e., either forward or reverse. For a new crystal, the first stress is automatically designated forward; the first stress reversal, then, is shown by a reverse arrow. An additional reverse of stress now is in the same direction as the initial one, so that the arrow again points forward. In the middle column, the total shear in radians is shown for the beginning of that sequence of runs. Thus, every new crystal starts with zero previous strain as shown. A negative sign indicates that the total reverse strain up to that time is greater than the total forward strain. The time interval between a run and a previous run is indicated by the fourth column, while the fifth gives the minimum load necessary to initiate creep on any particular sequence of runs.

Data are shown for three crystals, two of which (3A and 3B) are parts of one original crystal. The ability of a crystal, which has not been strained over too extreme a range, to anneal between runs taken

[21] F. Seitz, *Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943).

one day apart is shown by 3A. After considerable strain, however, the minimum stress necessary to initiate creep has risen from 28 to 35 grams per square millimeter.

Having established the fact that data taken on successive days were essentially reproducible and independent of previous runs, it was then possible to investigate in the other crystals the more detailed behavior of a crystal which is sheared alternately in one direction and then the other. A minimum load to cause the reverse creep was not determined for 3A, since the first load applied, unfortunately, was a 25 gram per square millimeter load, which caused rather rapid creep, showing that the minimum load was less then that and no more exact value could then be given.

3B exhibits the Bauschinger effect in a very definite manner. If 20 grams per square millimeter was sufficient to start the shear, then it is evident that the 11 needed in the reverse direction was so much lower as a result of the previous run. The effect appears to last an appreciable period of time, if very little strain is allowed to take place in the reverse direction, but after several runs have been made the effect appears to wear off completely. Rapid reversal enables runs to be made at extremely low stress values, and since there is a small amount of unaided creep recovery in any case, it is safe to assume that instantaneous reversal would find zero for the minimum load necessary to produce creep.

For crystal number 4, the same effect is again observed. However, the effect is the more pronounced in that hardening is simultaneously taking place, so that the crystal does not shear at the same load that it did previously, in the same direction. This hardening, which is not very apparent on the other crystals, will be discussed in detail later on.

Twinning

There are six equivalent twinning planes for an hexagonal close packed metal, all of the form $(10\bar{1}2)$.[13] For a zinc crystal, which has longitudinal to digonal axial ratio 1.86, the twinning plane will be parallel to a face of the specimen and to the shearing plates when the orientation is given by $\theta=43°$, $\alpha=30°$.

Three crystals (5, 6, 7) were grown with this orientation and mounted in the usual manner. Crystals of this orientation presented a difficult handling problem since they are extremely "soft" and since, further, if the Woods metal solder is not holding at every point on the surface, the orientation is such that the crystal will be easily pulled apart, as was the first crystal. In all three, however, it was found impossible to obtain any observable creep even up to the maximum loads which could be used in this apparatus. There was also no visible evidence of twinning having taken place. The third was sawed into two portions before being tested. After loading 7B, its x-ray back reflection pattern was compared with 7A, which had not been loaded at all. From the practically identical patterns at the points compared, by this means also, it was concluded that no twinning had occurred.

A recent article[14] quotes a figure of 300 to 450 grams per square millimeter as the stress along the twinning plane necessary to twin a zinc crystal. Since loads greater than 170 g/mm^2 were not obtained on this apparatus, the above results are not surprising.

Interest in twinning was originally stimulated by the observation (see Fig. 8) of creep in a plane perpendicular to the basal plane. It was thought that twinning might first have occurred, followed by creep along those basal planes which now would be in approximately the correct shearing direction. As a result of this portion of the investigation, it seems reasonable to conclude that creep along these prismatic planes is possible, although it has not previously been observed.

DISCUSSION

As might be expected, the orientation of the crystal in the apparatus strongly affects the results. In tensile experiments the orientation of the crystal makes little difference and only an easily computable factor is needed to obtain resolved shear stress from the actual stress (a function of θ and α). In our shear apparatus, as can be seen from the minimum stresses needed to initiate creep and from the dependence of the coefficient a on the stress σ, it is quite different. For θ as much as seven degrees away from the customary 90°, the yield point is extremely high (Table I, crystal 1), although creep occurs (Fig. 7).

Even for θ only two or three degrees away from the proper ninety degree orientation (crystal 4) there is a pronounced effect. The stress necessary to start the

TABLE II. The Bauschinger effect.

Crystal	Direction	Total Shear (radians)	Rest Interval	Yield Point (g/mm^2)
3A	→	0		28
		.025	1 Day	28
		.040	1 Day	28
		.100	5 Days	35
	←	.105	2 Days	<25
3B	→	0		22
	←	.016	2 Days	11
		.008	2 Days	17
		.002	1 Day	22
	→	-.004	20 Min.	6
		-.002	1 Day	17
4	→	0		29
		.0035	11 Hrs.	29
	←	.0038	15 Min.	2.7
		.0025	17 Hrs.	30
	→	-.0011	20 Min.	5.5
		.0004	11 Hrs.	33
		.0035	20 Min.	45
		.0093	3 Hrs.	52
	←	.0143	11 Hrs.	17
		.0110	36 Hrs.	41

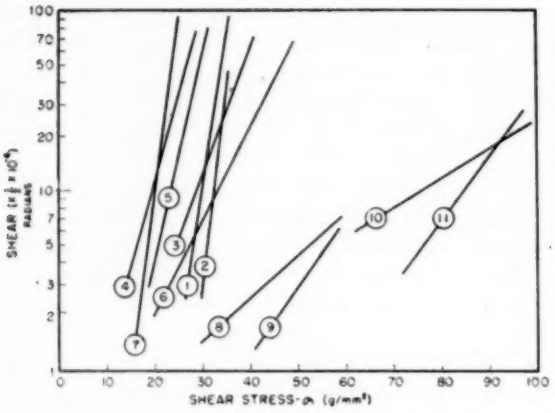


FIG. 13. The coefficient a *vs* the stress for various crystals under differing conditions as follows:

Line	Crystal	Description
1	3A	Three consecutive days of small strain runs
2	3A	After a total shear of about 3°
3	3A	Under reverse stress
4	3B	First forward runs
5	3B	First reversed runs (after 1 day)
6	8A	First forward runs
7	9A	First forward runs
8	4	First forward runs
9	4	First reversed runs
10	4	Second forward runs (note extreme hardening)
11	4	Second reversed runs

creep process appears to be about the same, but the rate of creep (i.e., as measured by a) is greatly reduced (Fig. 4 and Fig. 13). This implies that, as creep begins, the planes have to be rotated or bent somewhat for the process to continue, and it is not surprising then to see that the runs made on the following days exhibit a pronounced "hardening," since bent lattices undoubtedly block the motion of one shearing plane upon another (or, more properly, the motion of lattice imperfections along these planes). Also, the stress necessary to produce creep almost doubled from the initial run to one two days later. A similar effect is exhibited by crystal number 2 (Fig. 7). On the other hand, crystals 3 and 8, whose orientations proved upon cleaving to be within a half-degree of that desired, while exhibiting a definite hardening after considerable working, never required more than an additional 30–40 percent stress for creep to start. And, as is shown by crystal 3A (Fig. 2), the same loads and the same rates of creep were obtained three days in a row, showing that little time was required to anneal out the effects of the strain on previous runs.

The simplicity of the results presented here for the favorably oriented crystals even to fairly large strains (0.1 radian in Fig. 6) show the superiority of applying force to the crystal directly as a shear rather than relying on the resolution procedure of the tensile test method, as has indeed been noted by others. The validity of the tensile test method is particularly poor for anything but very small strains in the case where the testing machine is of such a nature that one end of the specimen is made to move away from the other along the direction of the initial long dimension of the specimen. Since the planes slip parallel to the glide direction (which must make an angle less than zero or greater than ninety degrees with the length of the specimen) a bending of the specimen and of the glide planes must occur in the neighborhood of the ends of the crystal and may extend appreciably into the part being measured. Such bent planes could not be expected subsequently to show simple glide. In the case where the upper end of the specimen is hung from a universal joint arrangement and a hanging load is applied by a second universal joint, it would seem that bending would not occur, but even here, if one region of a crystal has slipped appreciably more than neighboring regions above and below, the effect is to put a sort of kink in the crystal, and the straightening of this by the longitudinal force again causes bending. Moreover, in all cases where the ends are held fixed in mounting cups, simple glide cannot occur for planes which extend into these cups. In the case of large strains, where most previous data have been taken, the strains are never uniform over the length of the crystal, as evidenced by the development of slip bands and occasionally even twin regions over local areas on the surface of the crystal. This means that it cannot be expected that all lattice planes have glided by equal amounts, and hence, there will be considerable bending of one part of the crystal with respect to another. Evidently, this produces a rate of creep which is far from representative of that occurring at smaller strains and, further, a hardening which cannot readily be annealed out. Since much of the data of other observers have been obtained with thin wires stretched from five to several hundred percent, it is, therefore, not surprising if this internal bending effect should not contribute an appreciable amount to the discrepancies among reports from different observers. In this laboratory a crystal of zinc of the form used in the shearing apparatus was stretched in a tensile apparatus to a strain of some twenty percent. The eventual appearance of slip bands and large twinned sections was remarkable; it could not be doubted that a tremendous amount of internal bending and general lattice misalignment had taken place. This is the more significant since care had been taken to avoid any external nonaxial stresses by mounting the crystal in universal joints, both top and bottom, and adding the increments in an extremely careful and uniform manner. Thus, the contention of Orowan[25] that previous work taken at large strain regions could not be interpreted as meaningful data is supported, although the reasoning here is quite different. Orowan believes that strain hardening is so excessive at large strains that such procedures are valueless, whereas we believe that it is not actually strain hardening at all, but merely very nearly, or actually, macroscopic bending that is inherent in such large strain tensile measurements. As the long-term

[25] E. Orowan, Proc. Phys. Soc. (London) **52**, 8 (1940).

large-strain curve for crystal 9A (Fig. 6) shows, no such difficulty is encountered when the shear is applied in such a manner as to produce no lattice rotations. The face of the specimen indicated a uniform shear over the entire cross section, and no slip bands or twin regions were able to be detected.

The occurrence of even a single run which follows a single creep law from the inception of the run to a final strain of about 0.1 radian in a period of about five days should be considered extremely significant for, although many runs might be obtained which do not show such a simple relationship, in view of the above discussion, it is permissible to assume that some mechanism other than simple creep is taking place, either as a result of the method of growing or handling the specimen, or as a result of the way in which the loading or mounting was carried out. In other words, it would seem that one such run as that of Fig. 6 is sufficient to cast some doubt on the validity of relationships at variance with this simple law. Other laws are then the results of secondary processes which were not able to be eliminated by the methods employed.

Additional significant information can be obtained from the data available on the dependence of the coefficient a on the stress σ. Without regard to the particular form of this function, Figs. 11 and 12 show that in the stress region where creep can be obtained only by reversal of the stress direction, i.e., in the Bauschinger region, the dependence of a on σ is approximately a continuation of the same curve in the larger direct-stress regions. The left-hand plot of Fig. 12 furnishes the best proof of this statement, as can be seen from the fact that the open circle points, which were obtained from the initial run in the forward direction, establish a line along which the × points are seen to lie, although these latter were obtained after rapidly reversing the stress. Twelve hours later, as shown by the solid circles, continuation of the run in the same (reverse) direction, although requiring a higher initial stress, gave results which lie very well along the same line previously established.

The dashed line of Fig. 12 shows that the creep rate at any given stress is now lower than for the solid line. This is probably the result of the specimens having been, by this time, considerably strained, and this effect may be called a "hardening."

In the left-hand plot of Fig. 11 results similar to those of Fig. 12 are shown. The open circles again represent the initial run, with the × points again being those for the immediate reversal of the stress. Since the fluctuations are in all cases rather large, it is by no means certain that the × points and the open circles do belong to the same curve, but if one examines, for example, the solid line in Fig. 10, similar variations can be found where there was no reversal.

The solid circles in the left plot of Fig. 11 exhibit the same hardening as that of Fig. 12. These points again appear to lie approximately along a straight line which is displaced to the right with respect to the initial line, but again parallel to it. The triangle points, which were obtained about a day after the solid circles, but after reversing the stress again (i.e., the crystal is now being sheared in the initial forward direction), appear a little displaced from the line of solid circles, but well to the right of the initial open circles. This may be interpreted as hardening, again, though with a little less hardening having taken place for the forward direction than for the reverse.

Figure 10 also shows this hardening. Three days of runs at relatively low strains were obtained and are seen to lie along the solid line. The third day a large additional strain (up to about 0.1 radian) was obtained and the triangles, obtained several days later, exhibit a similar, more or less parallel displacement to the right. Several days later the specimen was strained under reverse stress, and the plusses form a line of different slope. No simple interpretation for this is offered, particularly in view of the large deformation accumulated previous to this sequence of runs.

Figure 13 is intended as a sort of summary sheet showing, in part, the same data as that of Figs. 10, 11, and 12 and for the remainder, data not previously exhibited in this manner. Lines 1, 2, 3 are, of course, from Fig. 10. Lines 4 and 5 are from data shown in Fig. 3. Here the points are sufficiently spread about these two lines so that one is probably not justified in attempting to explain the apparent displacement of the second line (5) from the first.

The outstanding conclusion to be drawn from Fig. 13 is that a specimen whose orientation is even slightly off from the proper one behaves in an obviously different manner from correctly oriented ones. This is clearly shown by lines 8, 9, 10, and 11, in which not only are the initial stresses to produce shear much higher, but the hardening after one set of runs was completed is extreme.

Some insight into one form of history dependence, namely, hardening, is offered by examination of Fig. 9. There data are shown for a normal creep run (open circles) and then for a run made at the same load value immediately after completion of the initial run (crosses). The steeper slope in the middle section of this latter curve is probably caused by the rapid annealing out of the strain hardening which causes the first portion of the curve to be nearly flat. When the effect of the previous run has been nearly annealed out, there follows the usual creep time relation, though the coefficient a never achieves the higher value which it had on the initial run. The slope m, however, becomes a constant of about the same value as that for the first run. The time required for this annealing to take place at room temperature seems reasonable because on previous runs spaced about 18 hours apart, no such hardening, or flattening of the curve, could be observed.

Briefly, then, the experimental data presented here show that a single creep law is valid over a wide range

of experimental conditions. Both forward and reversed stresses give rise to the same law, even in the extremely low stress region where creep can only be obtained by rapid reversal of the applied stress (i.e., in the Bauschinger region). Further, for the room temperature conditions employed throughout, it is evident that the exponent m remains constant over a wide stress and strain region with little history dependence, while the coefficient a retains the functional dependence of the strain on the stress, as well as on the previous history of the crystal.

If the plastic flow of metals has its origin in the multiplicative dislocation process, as is generally assumed today, then certain conclusions can properly be drawn from the above observations. In particular, those properties of a crystal specimen which might reasonably be expected to remain constant, with little regard for the previous history of the specimen, ought to be associated with m in the simple creep law here employed. Similarly, those properties which are variable with each specimen in an essential manner must then be considered as determining a.

Thus, since the number of dislocations at any instant of time is surely a variable dependent on the amount of shear, the time of anneal, the method of growing and handling, etc., we should expect to find this quantity incorporated into the coefficient a in no simple, or even single valued, relationship. On the other hand, any rate process theory, which might well apply to the migration of dislocations throughout the specimen, should predict the existence of such a constant as the exponent m, which should, however, be markedly dependent on the temperature. This temperature dependence is, indeed, already established[26] and shows the increase of m with the temperature in a pronounced manner.

The existence of an a different from zero in the low stress Bauschinger region, where no creep could initially be observed before reversing the stress, is merely demonstration of the production, by previous strain, of local stresses, which, together with the small applied stress, add up to the same approximate numerical stress value under which creep was observed before reversal. The number of these regions being presumably small compared with the number of regions taking part in the initial forward shearing process, the actual strain rate (as measured by a, m remaining constant) is also small as observations indicate.

The factor determining the number of regions of lattice imperfections, or dislocations, taking part may be assumed to be dependent on the blocking of the movement of the dislocations at any particular time rather than their actual creation or destruction, since otherwise one would not expect to find creep runs made on the same specimen on succeeding days to have almost identical values of a. One other possibility does exist, however, under which reproducible data could be obtained even though the flow process actually were to consist of the continual destruction (for m less than 1) of dislocations. An equilibrium number of dislocations might be found to exist at a particular temperature, and, if more or less are at some instant of time actually in the specimen, these would be expected with sufficient anneal time to reach the proper number for that temperature. Experimental verification of this latter possibility should be obtainable by annealing a specimen at some high temperature for sufficient time, then cooling it relatively rapidly, and shearing it at the lower temperature. The a determined in this manner might still show the effect of more than the equilibrium number of dislocations, i.e., the strain rate might be greater than normal, even though the value of m should be the same.

The author wishes to thank Professor E. P. T. Tyndall of the Physics Department of the State University of Iowa, whose time and effort in directing this work was generously given and whose critical evaluation at all stages was both fruitful and stimulating. In addition, the author wishes to express his appreciation for the cooperation shown by the entire solid state group and, in particular, by Mr. G. Alers whose willing assistance is gratefully remembered.

[26] D. O. Thompson (private communication).

Tolerance Coefficients for *R-C* Networks*

CHARLES BELOVE
Nuclear Development Associates, Inc., White Plains, New York
(Received December 20, 1952)

A method is presented for the calculation of the departure of actual network behavior from design characteristics attributable to the use of nonideal components. The method yields a set of tolerance coefficients relating percentage changes in the positions of the zeros and poles of the network function to percentage changes in the network components. Once this information is available, changes in gain or phase are easily determined.

1. INTRODUCTION

IN the practical design of networks we are often concerned with the departure of the actual network behavior from design characteristics caused by the use of nonideal components. The present paper is concerned with a method for determining these effects for certain *R–C* network functions. The results apply also to *R–L*, *L–C*, and analogous nonelectrical networks with only a change of symbols.

The method consists of the derivation of a set of tolerance coefficients relating percentage changes in the positions of the zeros and poles of the network function to percentage changes in the components of the network. Once this information is available, changes in gain or phase are easily determined. It turns out that an exact solution is tractable only when the network contains two or at most three independent capacitors. Two theorems will be proved which serve as checks on approximations made for more complex networks.

2. TOLERANCE COEFFICIENTS

Consider the principal determinant of the general *R–C* network. After clearing negative powers of p, the roots are found from an equation of the form

$$p^n\Delta = a_n p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0 = 0, \tag{1}$$

or in factored form,

$$(T_a p+1)(T_b p+1)\cdots = 0. \tag{2}$$

Here T_a is positive real for *R–C* networks and a homogeneous function of the network time constants. We have

$$T_a = f(T_{11}, T_{12}, T_{22}, \cdots), \tag{3}$$

where $T_{11}=R_1C_1$, $T_{12}=R_1C_2$, etc., and any R_i or C_i may be a series or parallel combination of several of the R's and C's of the network.

The total differential of T_a may be put in the form

$$\frac{dT_a}{T_a} = \left[\frac{T_{11}}{T_a}\frac{\partial T_a}{\partial T_{11}}\right]\frac{dT_{11}}{T_{11}} + \left[\frac{T_{12}}{T_a}\frac{\partial T_a}{\partial T_{12}}\right]\frac{dT_{12}}{T_{12}} + \cdots. \tag{4}$$

The quantities in brackets are called the tolerance coefficients. A given change in any T_{ij} when multiplied by the proper tolerance coefficient gives the resulting percentage change in T_a subject to the error in assuming that the differential is equal to the increment.

A very useful property of the tolerance coefficients is that their sum is exactly equal to one. To prove this, we note that attributable to the homogeneity of T_a, we have for arbitrary λ

$$f(\lambda T_{11}, \lambda T_{12}, \cdots) = \lambda f(T_{11}, T_{12}, \cdots), \qquad (5)$$
$$= \lambda T_a$$

Differentiating with respect to λ, setting $\lambda=1$, and dividing through by T_a yields

$$\frac{T_{11}}{T_a}\frac{\partial T_a}{\partial T_{11}} + \frac{T_{12}}{T_a}\frac{\partial T_a}{\partial T_{12}} + \cdots = 1, \tag{6}$$

where the left-hand side is the sum of the tolerance coefficients of (4).

If we substitute

$$\frac{dT_{ij}}{T_{ij}} = \frac{dR_i}{R_i} + \frac{dC_j}{C_j} \tag{7}$$

in Eq. (4) it is evident that we have theorem 1.

Theorem 1

The sum of the tolerance coefficients for the resistors is equal to one and similarly for the capacitors.

The theorem evidently holds for the roots of any of the symmetrical minors of Δ. It must also hold for the roots of the unsymmetrical minors even through they are not restricted to the negative real axis.

3. A PROPERTY OF THE ROOTS OF Δ

We next show that the reciprocals of the zeros of Δ for any *R–C* network obey relations of the form

$$\frac{\partial T_a}{\partial R_i} > 0 \text{ and } \frac{\partial T_a}{\partial C_i} > 0. \tag{8}$$

To establish this result we make use of the energy relations derived by Bode.[1] According to Bode's analysis

* This paper represents a portion of a thesis submitted to the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The work was supported by the U. S. Navy Bureau of Ordnance under contract NOrd 10740.

[1] H. W. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Company, Inc., New York, 1952), pp. 125–136.

the zeros of $\Delta(p)$ for an R–C network must obey a relation of the form

$$\left[F(a)+F(b)\right]+\frac{1}{p}\left[V(a)+V(b)\right]=0, \tag{9}$$

where $[F(a)+F(b)]$ corresponds to the "dissipation function" of the network and $[V(a)+V(b)]$ to the stored energy in the capacitors. These energy functions are positive definite quadratic forms and may be written

$$\left.\begin{aligned} F(a)&=\tfrac{1}{2}\sum_{r,s=1}^{r,s=n} R_{rs}I_{ra}I_{sa},\\ V(a)&=\tfrac{1}{2}\sum_{r,s=1}^{r,s=n}\frac{1}{C_{rs}}I_{ra}I_{sa}, \text{ etc,}\end{aligned}\right\} \tag{10}$$

where I_{ra} and I_{rb} are the real and imaginary parts of the transient mesh current I_r.

Suppose a zero occurs at $p=-\omega_a=-1/T_a$. Then

$$T_a=\frac{F(a)+F(b)}{V(a)+V(b)}. \tag{11}$$

If the network is planar and all mesh currents are taken in the same direction, it is evident that any particular resistance in the network, say R_k, can be made to appear in at most three of the terms of Eq. (11). If it is in a mutual branch R_{ij} it must appear also in R_{ii} and R_{jj} and if not it will appear only in R_{ii}. If the network is nonplanar, the currents can always be taken so that R_k appears in at most three of the terms. When this is done $F(a)$ may be written, noting that $R_{ij}=R_{ji}$,

$$\begin{aligned}2F(a)=\cdots&+(R_{ii}'+R_k)I_{ia}^2+2(R_{ij}'+R_k)I_{ia}I_{ja},\\ &+(R_{jj}'+R_k)I_{ja}^2+\cdots.\end{aligned} \tag{12}$$

Here $R_{ij}'=R_{ij}-R_k$ and the only terms which have been written down are those which involve R_k. Using Eqs. (11) and (12),

$$\frac{\partial T_a}{\partial R_k}=\frac{(I_{ia}+I_{ja})^2+(I_{ib}+I_{jb})^2}{V(a)+V(b)}. \tag{13}$$

Since this expression is always positive, Eq. (8) is proved for any R_k which appears in three of the terms of Eq. (12) and evidently holds for R_k in only one term as well.

For the capacitors, we consider the reciprocal of Eq. (11). An argument similar to Eq. (12) and (13) shows that

$$\frac{\partial \omega_a}{\partial C_k}<0, \tag{14}$$

and since $T_a=1/\omega_a$

$$\frac{\partial T_a}{\partial C_k}>0. \tag{15}$$

Thus Eq. (8) holds for any C_k and we have theorem 2.

Theorem 2

The zeros of Δ for any R–C network when expressed as time constants can only increase (decrease) if any resistor or capacitor in the network increases (decreases).

4. CONCLUSIONS

From the results of the previous sections, we may draw the following conclusions: (1) The magnitude of any zero of Δ for an R–C network, expressed as a time constant, can only increase if any resistor or capacitor in the network increases. (2) Since any symmetrical minor, Δ_{jj} may be thought of as the principal determinant of a network obtained from the original by eliminating the proper mesh, the above must be true for any Δ_{jj}. (3) For any of the above-mentioned time constants, the sum of the tolerance coefficients for the resistors must be equal to one and similarly for the capacitors. Noting this and (1) and (2) above the tolerance coefficients must all be positive and less than or equal to one. (4) For any network constructed using resistors all having the same fractional tolerance $\pm\delta$ and capacitors all having the same fractional tolerance, $\pm\epsilon$, the maximum shift in the zeros of Δ or Δ_{jj} from their design values will be $\pm(\delta+\epsilon)$. In particular this holds for both the zeros and poles of the driving point functions. (5) The only conclusion concerning the roots of the unsymmetrical minors Δ_{ij} is that the sums of the tolerace coefficients must be one.

5. COMPUTATION OF THE TOLERANCE COEFFICIENTS

Before the tolerance coefficients can be computed the network under consideration must be analyzed. Usually this yields the network function of interest in the form of a rational polynomial in which the coefficients are known functions of the R's and C's of the network. Obtaining the factored form (2) where the T's are known exactly in terms of the R's and C's is tractable only when the polynomials are of first, second, or at most third degree. For the second degree case we have

$$p^2+a_1p+a_0=(p+\omega_1)(p+\omega_2). \tag{16}$$

After some algebra,

$$\frac{d\omega_1}{\omega_1}=\frac{-a_1}{(a_1^2-4a_0)^{\frac{1}{2}}}\frac{da_1}{a_1}+\frac{2a_0}{[a_1(a_1^2-4a_0)^{\frac{1}{2}}-(a_1^2-4a_0)]}\frac{da_0}{a_0}, \tag{17}$$

$$\frac{d\omega_2}{\omega_2}=\frac{a_1}{(a_1^2-4a_0)^{\frac{1}{2}}}\frac{da_1}{a_1}-\frac{2a_0}{[a_1(a_1^2-4a_0)^{\frac{1}{2}}+(a_1^2-4a_0)]}\frac{da_0}{a_0}. \tag{18}$$

If design values have been chosen, numerical values for a_1 and a_0 should be substituted in the above and then da_0/a_0 and da_1/a_1 found in terms of dR_i/R_i and dC_j/C_j.

When higher order equations occur, Lin's method may be used to obtain literal approximations to the roots. Results will be fairly accurate if the roots are well separated.

If a driving point impedance or admittance is based directly on a partial fraction expansion, each pole will be determined by only one R and one C. The zeros are easily approximated in terms of the residue constants and poles since one zero must occur between each pair of poles.

6. SUMMARY

When the network under consideration has no more than two independent capacitors, the maximum degree of the determinantal equation is two and it is possible to compute all of the tolerance coefficients exactly using Eqs. (17) and (18). For more complex networks approximations must be made in order to obtain the zeros and poles in terms of the coefficients. The procedure is as follows:

(1) Analyze the network to obtain the network function in the form of a rational polynomial or partial fraction expansion if applicable. The coefficients in the first case and residue constants and poles in the second case must be obtained in terms of the R's and C's of the network.

(2) Compute the numerical values of the coefficients, zero, and poles.

(3) Obtain literal approximations to the zeros and poles using one of the methods described previously. Check these against the known numerical values.

(4) Derive the tolerance equations from the approximations found in 3.

(5) Check the sums, signs, and magnitudes of the tolerance coefficients noting Sec. 4.

The author wishes to express his gratitude to John L. Bower for his counsel and encouragement.

[2] Shih-Nge Lin, J. Math. and Phys. **XX**, 231, 1941. See also Brown and Campbell, *Principles of Servomechanisms* (John Wiley and Sons, Inc., New York, 1948), pp. 89–91.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Simplification for Mutual Impedance of Certain Antennas*

JESSE GERALD CHANEY†
United States Naval Postgraduate School, Monterey, California

(Received February 9, 1953)

The formula for mutual impedance by the generalized circuit method is reduced to a form which greatly reduces the number of integrations required in determining the mutual impedance of various combinations of open wire and terminated wire antennas. The application is illustrated by finding the mutual impedance of the legs of an open wire X-antenna, the latter being equivalent to the radiation impedance of a biconical antenna, and for small angle cones reduces to a form from which Schelkunoff's inverse terminal impedance may be determined.

GENERALIZED MUTUAL IMPEDANCE

THE mutual impedance of two antenna configurations is given in the generalized circuit by[1]

$$jkZ_{21}/30 = -\oint_1 \oint_2 Re[f_1(P_1)^* f_2(P_2)] \times \diamondsuit_1[e(r_{21})d\bar{r}_2]\cdot d\bar{r}_1, \quad (1)$$

in which

P_1 = any point along the axis of wire *one*,
P_2 = any point along the axis of wire *two*,
r_{21} = distance from P_1 to P_2,
$e(r_{21}) = r_{21}^{-1}\exp(-jkr_{21})$,
$k = \omega(\mu_0\epsilon_0)^{\frac{1}{2}} = 2\pi/\lambda$,
$\mu_0 = 4\pi(10)^{-7}$ *henries per meter*,
$\epsilon_0 = (36\pi\cdot 10^9)^{-1}$ *farads per meter*,
$\diamondsuit_1 = \nabla_1(\nabla_1\cdot) + k^2$ = operator *deltil*, with the subscript indicating the position at which the differentiations are to be performed,
$*$ = the complex conjugate to be taken,
$f_i(P_i)$ = the current distribution function along a wire,
Re = the real part to be taken.

With the exception of the distances involved in the case of two identical antennas, the integrations to be performed in Eq. (1) are the same as those to be performed in determining the self-radiation impedance by the conventional method of finding the mutual impedance between an axial path and a surface path. Equation (1) has been integrated for the self-radiation impedance of a terminated long wire antenna,[2] a terminated Vee-

* U. S. Naval Postgraduate School Technical Report No. 6, November, 1952.
† Professor of Electronics.
[1] J. G. Chaney, J. Appl. Phys. **22**, 12, 1429 (1951).

[2] J. G. Chaney, "On the generalized circuit theory as applied to antennas and radiating lines," U. S. Naval Postgraduate School Research Paper No. 1, March, 1951.

antenna,[3] and a terminated rhombic antenna.[3] Also, it has been integrated for the mutual impedance of two parallel staggered standing wave antennas of unequal length,[2] two parallel staggered terminated long wire antennas,[2] and for a standing wave antenna parallel staggered with a long wire antenna.[2]

In the case of the rhombic antenna, the algebra was greatly reduced by assuming the axial and surface paths to be interchangeable and by subjecting the currents to equations of constraint. This resulted in requiring the actual evaluation of only three integrals after postulating an unattenuated wave of current traveling along the rhombus. The resulting formula has been verified by the lengthy evaluation of the large number of integrals involved when the current distribution is postulated at the outset.[4] This reduction in algebra suggests that Eq. (1) be examined for similar reductions when applied in the derivation of formulas for mutual impedances.

Each of the two antennas (Fig. 1) will be assumed to be physically symmetric with respect to its generator. Hence,

$$jkZ_{21}/30=-\left(\oint_e^c+\oint_b^d\right)\left(\oint_f^n+\oint_m^g\right)\times Re[f_1(P_1)^*f_2(P_2)]\diamondsuit_1[e(r_{21})d\bar{r}_2]\cdot d\bar{r}_1. \quad (2)$$

Now let

$$A=-\left(\oint_e^c+\oint_b^d\right)\left(\oint_f^n+\oint_n^g\right)\times Re[f_1(P_1)^*f_2(P_2)]\nabla_1[\nabla_1\cdot e(r_{21})d\bar{r}_2]\cdot d\bar{r}_1. \quad (3)$$

Recall that

$$\nabla_1\cdot e(r_{21})d\bar{r}_2=-\nabla_2 e(r_{21})\cdot d\bar{r}_2 \quad (4)$$

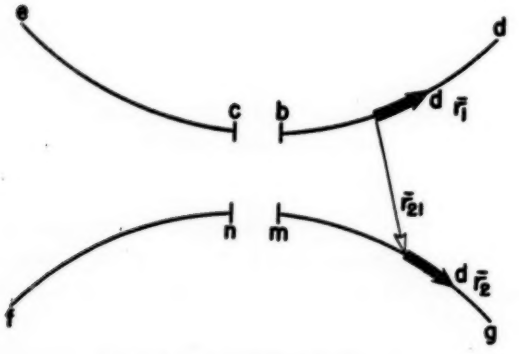


FIG. 1. Two coupled antennas.

and integrate by parts for

$$A=\left(\oint_f^n+\oint_m^g\right)\{Re[f_1(P_c)^*f_2(P_2)]\nabla_2 e(r_{2c})-Re[f_1(P_e)^*f_2(P_2)]\nabla_2 e(r_{2e})+Re[f_1(P_d)^*f_2(P_2)]\nabla_2 e(r_{2d})-Re[f_1(P_b)^*f_2(P_2)]\nabla_2 e(r_{2b})\}\cdot d\bar{r}_2+\left(\oint_e^c+\oint_b^d\right)\{e(r_{n1})\nabla_1 Re[f_2(P_n)^*f_1(P_1)]-e(r_{f1})\nabla_1 Re[f_2(P_f)^*f_1(P_1)]+e(r_{g1})\nabla_1 Re[f_2(P_g)^*f_1(P_1)]-e(r_{m1})\nabla_1 Re[f_2(P_m)^*f_1(P_1)]\}\cdot d\bar{r}_1+\left(\oint_e^c+\oint_b^d\right)\left(\oint_f^n+\oint_m^g\right)e(r_{21})\times\nabla_2\{\nabla_1 Re[f_1(P_1)^*f_2(P_2)]\cdot d\bar{r}_1\}\cdot d\bar{r}_2. \quad (5)$$

It is of interest to examine Eq. (5) for the three cases in which the two circuits are (1) both terminated rhombic antennas, (2) both open wire antennas, (3) one rhombic antenna and one open wire antenna.

In each case, not only are the antennas assumed to be symmetrically fed such that the currents are continuous through the generators, but the terminal spacings of the generators are assumed to be negligibly small in comparison with a wavelength. Likewise, the terminals for the load impedance of the rhombic antenna are assumed to be negligibly spaced. Then, since the current at the ends of a standing wave antenna vanish, the single integral terms of Eq. (5) vanish for each of the three cases. Thus,

$$A=\left(\oint_e^c+\oint_b^d\right)\left(\oint_f^n+\oint_m^g\right)\times e(r_{21})\nabla_2\{\nabla_1 Re[f_1(P_1)^*f_2(P_2)]\cdot d\bar{r}_1\}\cdot d\bar{r}_2. \quad (6)$$

Substituting into Eq. (1),

$$jkZ_{21}/30=\oint_1\oint_2 e(r_{21})[\nabla_2\{\nabla_1 Re[f_1(P_1)^*f_2(P_2)]\cdot d\bar{r}_1\}-k^2 Re\{f_1(P_1)^*f_2(P_2)\}d\bar{r}_1]\cdot d\bar{r}_2. \quad (7)$$

In terms of arc lengths along the antennas, let

$$g(ks_1, ks_2)=Re[f_1(P_1)^*f_2(P_2)]$$

$$\cos 2\alpha(s_1, s_2)ds_1 ds_2=d\bar{r}_1\cdot d\bar{r}_2,$$

[3] J. G. Chaney, "Free space radiation impedance of rhombic antenna," U. S. Naval Postgraduate School Technical Report No. 4, May, 1952.

[4] C. F. Klamm, Jr., "A detailed integration for the radiation impedance of a rhombic antenna," U. S. Naval Postgraduate School Technical Report No. 5, June, 1952.

and substitute into Eq. (7),

$$jkZ_{21}/30=\oint_1\oint_2 e(r_{21})\times\left[\frac{\partial^2}{\partial s_2\partial s_1}-k^2\cos2\alpha\right]g(ks_1, ks_2)ds_2ds_1. \quad (8)$$

Equation (8) is in a relatively simple form and is suitable for deriving formulas for the mutual impedances of several antenna configurations. It is true that any formula which requires an *a priori* assumption of the current distribution can be no more than an approximation. However, it has been found that many such formulas are sufficiently accurate for engineering purposes.

OPEN WIRE *X*-ANTENNA

An interesting application of Eq. (8) occurs in determining the mutual impedance of the legs of an open wire *X*-antenna (Fig. 2). The conventional sinusoidal current distribution will be assumed to be sufficiently accurate if the legs are not too near antiresonance. Hence,

$$g(ks_1, ks_2)=\sin k(l-|s_1|)\sin k(l-|s_2|)\csc^2kl, \quad (9a)$$

$$r_{12}^2=s_1^2+s_2^2-2s_1s_2\cos2\alpha, \quad (9b)$$

and

$$jkZ_{21}\sin^2kl/30=\int_{-l}^{+l}\int_{-l}^{+l}e(r_{21})\left[\frac{\partial^2}{\partial s_2\partial s_1}-k^2\cos2\alpha\right]\times\sin k(l-|s_1|)\sin k(l-|s_2|)ds_1ds_2. \quad (10)$$

The integrand has a singularity at the origin. However, one path may be integrated straight through the origin, whereas the other path may be assumed to have a gap of 2ϵ at the origin, and the limit taken as ϵ tends to zero. Thus,

$$\begin{aligned}jkZ_{21}\sin^2kl/30&=\int_{-l}^{0}\int_{-l}^{-\epsilon}e(r_{21})\left[\frac{\partial^2}{\partial s_1\partial s_2}-k^2\cos2\alpha\right]\\&\times\sin k(l+s_1)\sin k(l+s_2)ds_1ds_2\\&+\int_{-l}^{0}\int_{\epsilon}^{l}e(r_{21})\left[\frac{\partial^2}{\partial s_1\partial s_2}-k^2\cos2\alpha\right]\\&\times\sin k(l+s_1)\sin k(l-s_2)ds_1ds_2\\&+\int_{0}^{l}\int_{-l}^{-\epsilon}e(r_{21})\left[\frac{\partial^2}{\partial s_1\partial s_2}-k^2\cos2\alpha\right]\\&\times\sin k(l-s_1)\sin k(l+s_2)ds_1ds_2\\&+\int_{0}^{l}\int_{\epsilon}^{l}e(r_{21})\left[\frac{\partial^2}{\partial s_1\partial s_2}-k^2\cos2\alpha\right]\\&\times\sin k(l-s_1)\sin k(l-s_2)ds_1ds_2. \quad (11)\end{aligned}$$

Let

$$r_{13}^2=s_1^2+s_3^2+2s_1s_3\cos2\alpha, \quad (12)$$

and change the variables in Eq. (11),

$$\begin{aligned}jkZ_{21}\sin^2kl/60&=\int_0^l\int_\epsilon^l e(r_{21})\left[\frac{\partial^2}{\partial s_1\partial s_2}-k^2\cos2\alpha\right]\\&\times\sin k(l-s_1)\sin k(l-s_2)ds_1ds_2\\&-\int_0^l\int_0^l e(r_{13})\left[\frac{\partial^2}{\partial s_1\partial s_3}+k^2\cos2\alpha\right]\\&\times\sin k(l-s_1)\sin k(l-s_3)ds_1ds_3. \quad (13)\end{aligned}$$

After differentiating and transforming trigonometrically,

$$\begin{aligned}Z_{21}\sin^2kl/jk30&=\sin^2\alpha\exp(j2kl)\int_0^l\int_\epsilon^l r_{13}^{-1}\\&\times\exp[-jk(s_1+s_3+r_{13})]ds_1ds_3\\&+\sin^2\alpha\exp(-j2kl)\int_0^l\int_\epsilon^l r_{13}^{-1}\\&\times\exp[jk(s_1+s_3-r_{13})]ds_1ds_3\\&+\cos^2\alpha\int_0^l\int_\epsilon^l r_{13}^{-1}\exp[jk(s_1-s_3-r_{13})]ds_1ds_3\\&+\cos^2\alpha\int_0^l\int_\epsilon^l r_{13}^{-1}\\&\times\exp[-jk(s_1-s_3+r_{13})]ds_1ds_3\\&-\cos^2\alpha\exp(j2kl)\int_0^l\int_\epsilon^l r_{12}^{-1}\\&\times\exp[-jk(s_1+s_2+r_{12})]ds_1ds_2\\&-\cos^2\alpha\exp(-j2kl)\int_0^l\int_\epsilon^l r_{12}^{-1}\\&\times\exp[jk(s_1+s_2-r_{12})]ds_1ds_2\\&-\sin^2\alpha\int_0^l\int_\epsilon^l r_{12}^{-1}\exp[jk(s_1-s_2-r_{12})]ds_1ds_2\\&-\sin^2\alpha\int_0^l\int_\epsilon^l r_{12}^{-1}\\&\times\exp(-jk(s_1-s_2+r_{12})]ds_1ds_2. \quad (14)\end{aligned}$$

Upon comparing the definite integrals of Eq. (14) with the three integrals evaluated in deriving the self-

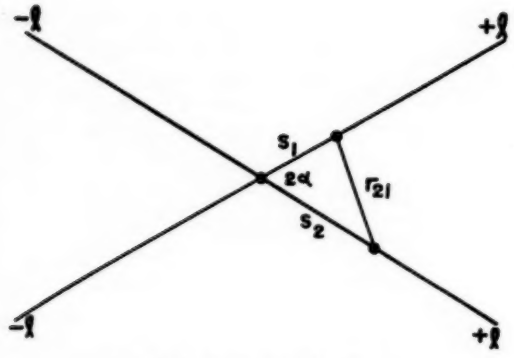


FIG. 2. An open wire X-antenna.

radiation impedance of a rhombic antenna,[3] it is found that the first, second, seventh, and eighth may be written directly, while the third, fourth, fifth, and sixth require only the replacement of the angle α by its complement. It might be mentioned that the value of the integrals do not depend upon the exact form in which the infinitesimal lower limit is taken.

Upon substituting into Eq. (14),

$$\begin{aligned}R_{21}\sin^2 kl/30 &= 2\ln\cot\alpha+2Ci(2kl\sin\alpha)-2Ci(2kl\cos\alpha)\\ &+\sin 2kl\{Si[2kl(1+\cos\alpha)]\\ &+Si[2kl(1-\cos\alpha)]-Si[2kl(1+\sin\alpha)]\\ &-Si[2kl(1-\sin\alpha)]\}+\cos 2kl\{2\ln\cot\alpha\\ &+Ci[(2kl(1+\cos\alpha)]+Ci[2kl(1-\cos\alpha)]\\ &-Ci[2kl(1+\sin\alpha)]-Ci[2kl(1-\sin\alpha)]\},\end{aligned}\quad (15a)$$

$$\begin{aligned}X_{21}\sin^2 kl/30 &= 2Si(2kl\cos\alpha)-2Si(2kl\sin\alpha)\\ &-\cos 2kl\{Si[2kl(1+\cos\alpha)\\ &-Si[2kl(1-\cos\alpha)]-Si[2kl(1+\sin\alpha)]\\ &+Si[2kl(1-\sin\alpha)]\}+\sin 2kl\{-4\ln\cot\alpha\\ &+Ci[2kl(1+\cos\alpha)]-Ci[2kl(1-\cos\alpha)]\\ &-Ci[2kl(1+\sin\alpha)]+Ci[2kl(1-\sin\alpha)]\}.\end{aligned}\quad (15b)$$

BICONICAL ANTENNA

Equations (15a, b) also give the radiation impedance of a biconical antenna under the sinusoidal current hypothesis, provided α is replaced by $\theta/2$, with 2θ being the apex angle of a nappe of the cone. The logarithmic term in Eq. (15b) becomes Schelkunoff's biconical characteristic impedance,[5]

$$K_0=120\ln\cot\theta/2. \quad (16)$$

By taking the apex angle sufficiently small for the current to be almost exactly sinusoidal, the driving point impedance of the biconical antenna may be evaluated from Eqs. (15a, b). Taking α very near zero and evaluating in the customary manner by writing

$$Si\epsilon\approx 0,\quad Ci\epsilon\approx\ln\gamma\epsilon=C+\ln\epsilon,\quad 0<\epsilon\ll 1, \quad (17)$$

where C is Euler's constant $0.5772\cdots$, the driving point impedance becomes,

$$\begin{aligned}Z_{in}=\{&60(C+\ln 2kl-Ci2kl)\\ &+30\sin 2kl(Si4kl-2Si2kl)\\ &+30\cos 2kl(C+\ln kl+Ci4kl-2Ci2kl)\\ &+j[60Si2kl+30\sin 2kl(Ci4kl-C-\ln kl\\ &-2\ln\cot\alpha)-30\cos 2klSi4kl]\}\csc^2 kl.\end{aligned}\quad (18)$$

After substituting Schelkunoff's inverse terminal impedance Z_a into Eq. (18), one obtains[6]

$$Z_{in}=Z_a\csc^2 kl-jK_0\cot kl, \quad (19)$$

which is the value used by Schelkunoff in the evaluation of his inverse terminal impedance, but which was arrived at by a slightly different procedure.[6]

---

[5] S. A. Schelkunoff, Proc. Inst. Radio Engrs. **29**, 9, 493 (1941).
[6] S. A. Schelkunoff, *Advanced Antenna Theory* (John Wiley and Sons, Inc., New York, 1952).

Mutual Impedance of Rhombic Antennas Spaced in Tandem*

JESSE GERALD CHANEY†
United States Naval Postgraduate School, Monterey, California
(Received February 9, 1953)

Upon examining the formulas for self- and mutual impedances of antennas, it is found that while the mutual impedance formula for separately driven collinear standing wave antennas may be used directly in determining the radiation impedance when such antennas are connected in cascade, certain modifications must be made in the case of traveling wave antennas under similar circumstances. Accordingly, formulas are derived for rhombic antennas spaced in tandem and are modified to permit the determining of the radiation impedance of two identical rhombic antennas connected in cascade.

IN TANDEM *VS* IN CASCADE

THE mesh equations for two coupled antennas are

$$\begin{aligned} Z_{11}I_1+Z_{12}I_2&=V_1,\\ Z_{21}I_1+Z_{22}I_2&=V_2,\\ Z_{12}&=Z_{21}. \end{aligned} \tag{1}$$

The radiation impedance parameters are given by certain double line integrals,[1] in which both paths are along the same wire for a self-impedance, whereas one path is along each wire for a mutual impedance.

In the case of an open wire full wave antenna fed at a current antinode, a conventional method for determining the driving point impedance is to set

$$V_2=0, \quad I_2=-I_1 \tag{2}$$

and add the two equations, giving

$$Z_{in}=2(Z_{11}-Z_{12}). \tag{3}$$

The validity of the above procedure rests in the fact that not only are the integration paths for Eq. (3) exactly the same as they would be if the paths were directly set up for the self-impedance of the full wave antenna, but the assumed currents satisfy the restriction

$$I_1^*I_2=Re(I_1^*I_2)=-|I_1|^2. \tag{4}$$

On the other hand, if the reference currents were not in time phase and Eq. (4) were not satisfied, say

$$I_2=I_1\exp(jkh), \tag{5}$$

the division by $|I_1|^2$, as required in the formulation of a self-impedance from the expression for the complex power, instead of by $I_1^*I_2$ as required for the formulation of the mutual impedance, would require the current factor,[1] $Re[f_1(P_1)^*f_2(P_2)]$, to be replaced by the factor,

$$Re[f_1(P_1)^*f_2(P_2)\exp(jkh)].$$

In other words, a mutual impedance formula for two individually driven travelling wave antennas would require modification before it would be permissible to write the radiation impedance of two such identical antennas driven in cascade by the formula

$$Z_r=2(Z_{11}+Z_{12}). \tag{6}$$

RHOMBIC ANTENNAS IN TANDEM

It has been shown[2] that the mutual impedance of two terminated rhombic antennas is given by

$$jkZ_{12/30}=\oint_1\oint_2 e(r_{12})\times\left[\frac{\partial^2}{\partial s_1\partial s_2}-k^2\cos\theta(s_1,s_2)\right]g(ks_1,ks_2)ds_1ds_2, \tag{7}$$

in which

$g(ks_1, ks_2)=Re[f_1(ks_1)^*f_2(ks_2)]$,
$e(r_{12})=r_{12}^{-1}\exp(-jkr_{12})$,
$f_1(ks_1)=$ *current distribution function along antenna one*,
$f_2(ks_2)=$ *current distribution function along antenna two*,
$\theta(s_1, s_2)=$ *angle between directions along the two antennas*,
$r_{12}=$ *distance between points on the two antennas*,
$k=2\pi/\lambda$.

For two identical, coaxial, and coplanar rhombic antennas, the same corresponding sides are parallel as in the case of the paths which were used for finding the self-radiation impedance of a rhombic antenna.[3] Since the current distribution functions are the same as those for the two paths along the same antenna, it follows that the integrations in Eq. (7) need only to be carried out over the nonparallel wires.[3]

Let each leg of the rhombus be of length l, and let the vertex angles at the generator be 2α. Then, after postulating unattenuated traveling waves of current

* U. S. Naval Postgraduate School Technical Report No. 7, December, 1952.
† Professor of Electronics.
[1] J. G. Chaney, J. Appl. Phys. **22**, 1429 (1951).
[2] J. G. Chaney, "Simplification for mutual impedance of certain antennas," U. S. Naval Postgrad. School Tech. Rpt. No. 6 (November, 1952).
[3] J. G. Chaney, "Free space radiation impedance of a rhombic antenna," U. S. Naval Postgrad. School Tech. Rpt. No. 4 (May, 1952).

along each rhombic antenna, Eq. (7) becomes

$$Z_{12}=j120\sin^2\alpha\Big[\int_0^l\int_0^l \cos k(x_1-x_2)e(r_{12})dx_1dx_2 + \int_0^l\int_0^l \cos k(x_1-x_3)e(r_{13})dx_1dx_3 - \int_0^l\int_0^l \cos k(x_1+x_4)e(r_{14})dx_1dx_4 - \int_0^l\int_0^l \cos k(x_1+x_5)e(r_{15})dx_1dx_5\Big], \quad (8)$$

or changing to the exponential form,

$$Z_{12}=-j60\sin^2\alpha[I_1+I_2+I_3+I_4-2I_5-2I_7], \quad (9)$$

in which

$$I_1=\int_0^l\int_0^l \exp[-jk(x_5+x_1+r_{15})]r_{15}^{-1}dx_5dx_1, \quad (10)$$

$$I_2=\int_0^l\int_0^l \exp[jk(x_5+x_1-r_{15})]r_{15}^{-1}dx_5dx_1, \quad (11)$$

$$I_3=\int_0^l\int_0^l \exp[-jk(x_4+x_1+r_{14})]_{14}^{-1}dx_4dx_1, \quad (12)$$

$$I_4=\int_0^l\int_0^l \exp[jk(x_4+x_1-r_{14})]r_{14}^{-1}dx_4dx_1, \quad (13)$$

$$I_5=\int_0^l\int_0^l \exp[jk(x_3-x_1-r_{13})]r_{13}^{-1}dx_3dx_1, \quad (14)$$

$$I_7=\int_0^l\int_0^l \exp[jk(x_2-x_1-r_{12})]r_{12}^{-1}dx_2dx_1, \quad (15)$$

in which x_0 is the distance between the driving points, and in which

$$r_{12}=[x_1^2+x_2^2+x_0^2-2x_0(x_1-x_2)\times\cos\alpha-2x_1x_2\cos2\alpha]^{\frac{1}{2}}, \quad (16)$$

$$r_{13}=[x_1^2+x_3^2+x_0^2+2x_0(x_1-x_3)\times\cos\alpha-2x_1x_3\cos2\alpha]^{\frac{1}{2}}, \quad (17)$$

$$r_{14}=[x_1^2+x_5^2+x_0^2+2x_0^2(x_1+x_4)\times\cos\alpha+2x_1x_4\cos2\alpha]^{\frac{1}{2}}, \quad (18)$$

$$r_{15}=[x_1^2+x_5^2+x_0^2-2x_0(x_1+x_5)\times\cos\alpha+2x_1x_5\cos2\alpha]^{\frac{1}{2}}. \quad (19)$$

Let

$$d_1=[x_0^2+(2l\sin\alpha)^2]^{\frac{1}{2}},$$
$$d_2=[x_0^2+l^2-2lx_0\cos\alpha]^{\frac{1}{2}},$$
$$d_3=[x_0^2+l^2+2lx_0\cos\alpha]^{\frac{1}{2}},$$

and

$$Z_{12}=R_{12}+jX_{12}.$$

After evaluating the definite integrals in Eq. (9), the following formulas are obtained:

$$\begin{aligned}
R_{12}/60=&\cos k(x_0\sec\alpha)\{4Cik(x_0\sec\alpha+l+d_3)+4Cik(x_0\sec\alpha+l-d_3)+4Cik(x_0\sec\alpha-l+d_2)\\
&+4Cik(x_0\sec\alpha-l-d_2)-4Cik[(\sec\alpha+1)x_0]-4Cik[x_0(\sec\alpha-1)]-2Cik(x_0\sec\alpha+d_1)\\
&-2Cik(x_0\sec\alpha-d_1)-Cik[x_0(\sec\alpha-1)+2l(1-\cos\alpha)]-Cik[x_0(\sec\alpha-1)-2l(1-\cos\alpha)]\\
&-Cik[x_0(\sec\alpha+1)+2l(1+\cos\alpha)]-Cik[x_0(\sec\alpha+1)-2l(1+\cos\alpha)]\}\\
&+\sin k(x_0\sec\alpha)\{4Sik(x_0\sec\alpha+l+d_3)+4Sik(x_0\sec\alpha+l-d_3)+4Sik(x_0\sec\alpha-l+d_2)\\
&+4Sik(x_0\sec\alpha-l-d_2)-4Sik[x_0(\sec\alpha-1)]-4Sik[x_0(\sec\alpha+1)]-2Sik(x_0\sec\alpha+d_1)\\
&-2Sik(x_0\sec\alpha-d_1)-Sik[x_0(\sec\alpha+1)+2l(1+\cos\alpha)]-Sik[x_0(\sec\alpha+1)-2l(1+\cos\alpha)]\\
&-Sik[x_0(\sec\alpha-1)+2l(1-\cos\alpha)]-Sik[x_0(\sec\alpha-1)-2l(1-\cos\alpha)]\}\\
&-\cos k(x_0\cos\alpha)\{2Cik(d_3+l+x_0\cos\alpha)+2Cik(d_3-l-x_0\cos\alpha)+2Cik(d_2+l-x_0\cos\alpha)\\
&+2Cik(d_2-l+x_0\cos\alpha)-4Cik[x_0(1+\cos\alpha)]-4Cik[x_0(1-\cos\alpha)]\}\\
&+\sin k(x_0\cos\alpha)\{2Sik(d_3+l+x_0\cos\alpha)+2Sik(d_3-l-x_0\cos\alpha)+2Sik(d_2+l-x_0\cos\alpha)\\
&+2Sik(d_2-l+x_0\cos\alpha)-4Sik[x_0(1+\cos\alpha)]-4Sik[x_0(1-\cos\alpha)]\}\\
&+\cos k(x_0\cos\alpha+2l\sin^2\alpha)\{Cik(d_1+x_0\cos\alpha+2l\sin^2\alpha)+Cik(d_1-x_0\cos\alpha-2l\sin^2\alpha)\\
&-2Cik(d_2+x_0\cos\alpha-l\cos2\alpha)-2Cik(d_2-x_0\cos\alpha+l\cos2\alpha)+Cik[(x_0-2l\cos\alpha)(1+\cos\alpha)]\\
&+Cik[(x_0-2l\cos\alpha)(1-\cos\alpha)]\}\\
&+\sin k(x_0\cos\alpha+2l\sin^2\alpha)\{Sik(d_1+x_0\cos\alpha+2l\sin^2\alpha)-Sik(d_1-x_0\cos\alpha-2l\sin^2\alpha)\\
&-2Sik(d_2+x_0\cos\alpha-l\cos2\alpha)+2Sik(d_2-x_0\cos\alpha+l\cos2\alpha)+Sik[(x_0\cos\alpha-l\cos2\alpha)(1+\cos\alpha)]\\
&-Sik[(x_0-2l\cos\alpha)(1-\cos\alpha)]\}\\
&+\cos k(x_0\cos\alpha-2l\sin^2\alpha)\{Cik(d_1+x_0\cos\alpha-2l\sin^2\alpha)+Cik(d_1-x_0\cos\alpha+2l\sin^2\alpha)\\
&-2Cik(d_3+x_0\cos\alpha+l\cos2\alpha)-2Cik(d_3-x_0\cos\alpha-l\cos2\alpha)+Cik[(x_0+2l\cos\alpha)(1+\cos\alpha)]\\
&+Cik[(x_0+2l\cos\alpha)(1-\cos\alpha)]\}\\
&+\sin k(x_0\cos\alpha-2l\sin^2\alpha)\{Sik(d_1+x_0\cos\alpha-2l\sin^2\alpha)-2Sik(d_3+x_0\cos\alpha+l\cos2\alpha)\\
&-Sik(d_1-x_0\cos\alpha+2l\sin^2\alpha)+2Sik(d_3-x_0\cos\alpha-l\cos2\alpha)+Sik[(x_0+2l\cos\alpha)(1+\cos\alpha)]\\
&-Sik[(x_0+2l\cos\alpha)(1-\cos\alpha)]\}, \quad (20)
\end{aligned}$$

$$
\begin{aligned}
X_{12}=&\cos k(x_0\sec\alpha)\{4Sik(x_0\sec\alpha+l-d_3)-4Sik(x_0\sec\alpha+l+d_3)-4Sik(x_0\sec\alpha+l-d_2)\\
&+4Sik(x_0\sec\alpha-l-d_2)+4Sik[x_0(\sec\alpha+1)]-4Sik[x_0(\sec\alpha-1)]+2Sik(x_0\sec\alpha+d_1)\\
&-2Sik(x_0\sec\alpha-d_1)+Sik[x_0(\sec\alpha+1)-2l(1+\cos\alpha)]+Sik[x_0(\sec\alpha+1)+2l(1+\cos\alpha)]\\
&-Sik[x_0(\sec\alpha-1)-2l(1-\cos\alpha)]-Sik[x_0(\sec\alpha-1)+2l(1-\cos\alpha)]\}\\
&-\sin k(x_0\sec\alpha)\{4Cik(x_0\sec\alpha+l-d_3)-4Cik(x_0\sec\alpha+l+d_3)-4Cik(x_0\sec\alpha-l+d_2)\\
&+4Cik(x_0\sec\alpha-l-d_2)+4Cik[x_0(\sec\alpha-1)]-4Cik[x_0(\sec\alpha-1)]+2Cik(x_0\sec\alpha+d_1)\\
&-2Cik(x_0\sec\alpha-d_1)+Cik[x_0(\sec\alpha+1)-2l(1+\cos\alpha)]+Cik[x_0(\sec\alpha+1)+2l(1+\cos\alpha)]\\
&-Cik[x_0(\sec\alpha-1)-2l(1-\cos\alpha)]-Cik[x_0\ (\sec\alpha-1)+2l(1-\cos\alpha)]\}\\
&+\cos k(x_0\cos\alpha)\{2Sik(d_3-l-x_0\cos\alpha)+2Sik(d_3+l+x_0\cos\alpha)+2Sik(d_2-l+x_0\cos\alpha)\\
&+2Sik(d_2+l-x_0\cos\alpha)-4Sik[x_0(1+\cos\alpha)]-4Sik[x_0(1-\cos\alpha)]\}\\
&+\sin k(x_0\cos\alpha)\{2Cik(d_3-l-x_0\cos\alpha)-2Cik(d_3+l+x_0\cos\alpha)-2Cik(d_1-l+x_0\cos\alpha)\\
&+2Cik(d_1+l-x_0\cos\alpha)+4Cik[x_0(1+\cos\alpha)]-4Cik[x_0(1-\cos\alpha)]\}\\
&-\cos k(x_0\cos\alpha+2l\sin^2\alpha)\{Sik(d_1+x_0\cos\alpha+2l\sin^2\alpha)+Sik(d_1-x_0\cos\alpha-2l\sin^2\alpha)\\
&-2Sik(d_2+x_0\cos\alpha-l\cos 2\alpha)-2Sik(d_2-x_0\cos\alpha+l\cos 2\alpha)+Sik[(x_0-2l\cos\alpha)(1+\cos\alpha)]\\
&+Sik[(x_0-2l\cos\alpha)(1-\cos\alpha)]\}\\
&+\sin k(x_0\cos\alpha+2l\sin^2\alpha)\{Cik(d_1+x_0\cos\alpha+2l\sin^2\alpha)-Cik(d_1-x_0\cos\alpha-2l\sin^2\alpha)\\
&-2Cik(d_2+x_0\cos\alpha-l\cos 2\alpha)+2Cik(d_2-x_0\cos\alpha+l\cos 2\alpha)+Cik[(x_0-2l\cos\alpha)(1+\cos\alpha)]\\
&-Cik[(x_0-2l\cos\alpha)(1-\cos\alpha)]\}\\
&-\cos k(x_0\cos\alpha-2l\sin^2\alpha)\{Sik(d_1+x_0\cos\alpha-2l\sin^2\alpha)+Sik(d_1-x_0\cos\alpha+2l\sin^2\alpha)\\
&-2Sik(d_3+x_0\cos\alpha+l\cos 2\alpha)-2Sik(d_3-x_0\cos\alpha-l\cos 2\alpha)+Sik[(x_0+2l\cos\alpha)(1+\cos\alpha)]\\
&+Sik[(x_0+2l\cos)(1-\cos\alpha)]\}\\
&+\sin k(x_0\cos\alpha-2l\sin^2\alpha)\{Cik(d_1+x_0\cos\alpha-2l\sin^2\alpha)-Cik(d_1-x_0\cos\alpha+2l\sin^2\alpha)\\
&-2Cik(d_3+x_0\cos\alpha+l\cos 2\alpha)+2Cik(d_3-x_0\cos\alpha-l\cos 2\alpha)+Cik[(x_0+2l\cos\alpha)(1+\cos\alpha)]\\
&-Cik[(x_0+2l\cos\alpha)(1-\cos\alpha)]\}. \qquad (21)
\end{aligned}
$$

For two closely spaced identical rhombic antennas, Eqs. (20) and (21) may be considerably simplified. At the frequency for which a rhombic antenna is given an optimum design, the leg length is almost always taken as an integral multiple of a half-wavelength with the parameter $2kl$ becoming an even multiple of π. Under these conditions, the formulas may be further simplified. However, when operating at a frequency different from the design frequency, $2kl$ is usually not an integral multiple of π, and the formulas must be more carefully examined in case it is desired to connect the antennas in cascade. Accordingly, select

$$x_0=2l\cos\alpha+b, \quad b\ll 2l\cos\alpha, \quad b\ll l, \qquad (22)$$

let

$$d_4=l(1+8\cos^2\alpha)^{\frac{1}{2}},$$

and simplify:

$$
\begin{aligned}
R_{12}/60=&\cos 2kl\{C+\ln(kl\sin^2\alpha)+2Ci2kl-Cik4kl-4Ci[2kl(1+\cos\alpha)]-4Cik[2kl(1-\cos\alpha)]\\
&-Ci[4kl(1+\cos\alpha)]-Ci[4kl(1-\cos\alpha)]+4Cik(3l+d_4)+4Cik(3l-d_4)\}\\
&+\sin 2kl\{2Si2kl-Si4kl-4Si[2kl(1+\cos\alpha)]-4Si[2kl(1-\cos\alpha)]-Si[4kl(1+\cos\alpha)]\\
&-Si[4kl(1-\cos\alpha)]+4Sik(3l+d_4)+4Sik(3l-d_4)\}\\
&-2\cos(2kl\cos^2\alpha)\{Ci(2kl\sin^2\alpha)+Ci(2kl\cos^2\alpha)-2Ci[2kl\cos\alpha(1+\cos\alpha)]\\
&-2Ci[2kl\cos\alpha(1-\cos\alpha)]+Cik(d_4+2l\cos^2\alpha)+Cik(d_4-2l\cos^2\alpha)\}\\
&+2\sin(2kl\cos^2\alpha)\{Si(2kl\sin^2\alpha)+Si(2kl\cos^2\alpha)-2Si[(2kl\cos\alpha)(1+\cos\alpha)]\\
&-2Si[2kl\cos\alpha(1-\cos\alpha)]+Sik(d_4+2l\cos^2\alpha)+Sik(d_4-2l\cos^2\alpha)\}\\
&+\cos(2kl\cos 2\alpha)\{Ci(4kl\cos^2\alpha)+Ci(4kl\sin^2\alpha)+Cik[4kl\cos\alpha(1+\cos\alpha)]\\
&+Cik[4kl\cos\alpha(1-\cos\alpha)]-2Cik(d_4+4l\cos^2\alpha-l)-2Cik(d_4-4l\cos^2\alpha+l)\}\\
&+\sin(2kl\cos 2\alpha)\{Si(4kl\cos^2\alpha)-Si(4kl\sin^2\alpha)+Sik[4l\cos\alpha(1+\cos\alpha)]\\
&-Sik[4l\cos\alpha(1-\cos\alpha)]-2Sik(d_4+4l\cos^2\alpha-l)+2Sik(d_4-4l\cos^2\alpha+l)\}, \qquad (23)
\end{aligned}
$$

$$X_{12}=\cos 2kl\{-2Si2kl+Si4kl+4Si[2kl(1+\cos\alpha)]-4Si[2kl(1-\cos\alpha)]+Si[4kl(1+\cos\alpha)]$$
$$-Si[4kl(1-\cos\alpha)]-4Sik(3l+d_4)+4Sik(3l-d_4)\}$$
$$-\sin 2kl\{C+\ln(kl\tan^2\alpha)-2Ci2kl+Ci4kl+4Ci[2kl(1+\cos\alpha)]-4Cik[2kl(1-\cos\alpha)]$$
$$+Ci[4kl(1+\cos\alpha)]-Ci[4kl(1-\cos\alpha)]-4Cik(3l+d_4)+4Cik(3l-d_4)\}$$
$$+2\cos(2kl\cos^2\alpha)\{Sik(2l\sin^2\alpha)+Si(2kl\cos^2\alpha)-2Si[2kl\cos\alpha(1+\cos\alpha)]-2Si[2kl\cos\alpha(1-\cos\alpha)]$$
$$+Sik(d_4+2l\cos^2\alpha)+Sik(d_4-2l\cos^2\alpha)\}$$
$$+2\sin(2kl\cos^2\alpha)\{Ci(2kl\sin^2\alpha)-Ci(2kl\cos^2\alpha)+2Ci[2kl\cos\alpha(1+\cos\alpha)]-2Ci[2kl\cos\alpha(1-\cos\alpha)]$$
$$-Cik(d_4+2l\cos^2\alpha)+Cik(d_4-2l\cos^2\alpha)]\}$$
$$-\cos(2kl\cos 2\alpha)\{Si(4kl\sin^2\alpha)+Si(4kl\cos^2\alpha)+Si[4kl\cos\alpha(1+\cos\alpha)]+Si[4kl\cos\alpha(1-\cos\alpha)]$$
$$-2Sik(d_4+4l\cos^2\alpha-l)-2Sik(d_4-4l\cos^2\alpha+l)\}$$
$$+\sin(2kl\cos 2\alpha)\{-Ci(4kl\sin^2\alpha)+Ci(4kl\cos^2\alpha)+Ci[4kl\cos\alpha(1+\cos\alpha)]-Ci[4kl\cos\alpha(1-\cos\alpha)]$$
$$-2Cik(d_4+4l\cos^2\alpha-l)+2Cik(d_4-4l\cos^2\alpha+l)\}, \tag{24}$$

in which $C=0.5772\cdots$ *is Euler's constant.*

Upon selecting the leg length as an integral multiple of a half-wavelength, formulas (23) and (24) reduce to

$$R_{12}/60=C+\ln(kl\sin^2\alpha)-Ci4kl+2Ci2kl-4Ci[2kl(1+\cos\alpha)]-4Ci[2kl(1-\cos\alpha)]-Ci[4kl(1+\cos\alpha)]$$
$$-Ci[4kl(1-\cos\alpha)]+4Cik(3l+d_4)+4Ci(3l-d_4)\}$$
$$-2\cos(2kl\sin^2\alpha\{Ci(2kl\sin^2\alpha)+Ci(2kl\cos^2\alpha)-2Ci[2kl\cos\alpha(1+\cos\alpha)]-2Ci[2kl\cos\alpha(1-\cos\alpha)]$$
$$+Cik(d_4+2l\cos^2\alpha)+Cik(d_4-2l\cos^2\alpha)]\}$$
$$-2\sin(2kl\sin^2\alpha)\{Sik(2kl\sin^2\alpha)+Si(2kl\cos^2\alpha)-2Si[2kl\cos\alpha(1+\cos\alpha)]$$
$$-2Si[2kl\cos\alpha(1-\cos\alpha)]+Sik(d_4+2l\cos^2\alpha)+Sik(d_4-2l\cos^2\alpha)\}$$
$$-\cos(4kl\sin^2\alpha)\{-Ci(4kl\sin^2\alpha)-Ci(4kl\cos^2\alpha)-Ci[4kl\cos\alpha(1+\cos\alpha)]-Ci[4kl\cos\alpha(1-\cos\alpha)]$$
$$+2Cik(d_4+4l\cos^2\alpha-l)+2Cik(d_4-4l\cos^2\alpha+l)\}$$
$$-\sin(4kl\sin^2\alpha)\{-Si(4kl\sin^2\alpha)+Si(4kl\cos^2\alpha)+Si[4kl\cos\alpha(1+\cos\alpha)]-Si[4kl\cos\alpha(1-\cos\alpha)]$$
$$+2Sik(d_4-4l\cos^2\alpha+l)-2Si(d_4+4l\cos^2\alpha-l)\}, \tag{25}$$

$$X_{12}/60=Si4kl-2Si2kl+4Si[2kl(1+\cos\alpha)]-4Si[2kl(1-\cos\alpha)]+Si[4kl(1+\cos\alpha)]-Si[4kl(1-\cos\alpha)]$$
$$+4Sik(3l-d_4)-4Sik(3l+d_4)$$
$$+2\cos(2kl\sin^2\alpha)\{Si(2kl\sin^2\alpha)+Si(2kl\cos^2\alpha)-2Si[2kl\cos\alpha(1+\cos\alpha)]-2Si[2kl\cos\alpha(1-\cos\alpha)]$$
$$+Sik(d_4+2l\cos^2\alpha)+Sik(d_4-2l\cos^2\alpha)\}$$
$$-2\sin(2kl\sin^2\alpha)\{Ci(2kl\sin^2\alpha)-Ci(2kl\cos^2\alpha)+2Ci[2kl\cos\alpha(1+\cos\alpha)]-2Ci][2kl\cos\alpha(1-\cos\alpha)]$$
$$-Cik(d_4+2l\cos^2\alpha)+Ci(d_4-2l\cos^2\alpha)\}$$
$$+\cos(4kl\sin^2\alpha)\{-Si(4kl\sin^2\alpha)-Si(4kl\cos^2\alpha)-Si[4kl\cos\alpha(1+\cos\alpha)]-Si[4kl\cos\alpha(1-\cos\alpha)]$$
$$+2Sik(d_4+4l\cos^2\alpha-l)+2Sik(d_4-4l\cos^2\alpha+l)\}$$
$$-\sin(4kl\sin^2\alpha)\{-Ci(4kl\sin^2\alpha)+Ci(4kl\cos^2\alpha)+Ci[4kl\cos\alpha(1+\cos\alpha)]-Ci[4kl\cos\alpha(1-\cos\alpha)]$$
$$+2Cik(d_4-4l\cos^2\alpha+l)-2Cik(d_4+4l\cos^2\alpha-l)\}. \tag{26}$$

Formulas (20) and (21) are for any two rhombic antennas that are terminated, have equal leg lengths, are coplanar, and are coaxially spaced x_0 between driving points. Formulas (23) and (24) are for any two identical rhombic antennas which are closely spaced in tandem. Formulas (25) and (26) are for two identical rhombic antennas whose leg lengths are an integral multiple of a half-wavelength and which are closely spaced in tandem. It will be shown that formulas (25) and (26) may also be used for rhombic antennas connected in cascade even though the leg lengths are arbitrary.

RHOMBIC ANTENNAS IN CASCADE

If instead of solving for the phase of I_2 as in the case of separately driven antennas, the phase of I_2 is to be specified with respect to I_1 in accordance with Eq. (5), an examination of the integrations of Eqs. (10) to (15), inclusive, reveals that formulas (20) and (21) require modification only in the sine and cosine coefficients of the sine integral and cosine integral functions within the braces. This modification consists merely of adding the term kh to the argument of the sine and cosine coefficients. For example, $\cos k(x_0\sec\alpha)$ is replaced by $\cos k(h+x_0\sec\alpha)$. The algebraic sign of h is always the same as that of x_0.

If two identical rhombic antennas are to be connected in cascade, in order that a form of Z_{12} suitable for substitution into Eq. (6) may be written, select $h=-2l$ and modify formulas (23) and (24). This again gives the formulas (25) and (26). Thus the particular form given for the mutual impedance of two closely spaced rhombic

antennas in formulas (25) and (26) is also valid for antennas connected in cascade even though the leg lengths are not an even multiple of a half-wavelength.

Hence, it follows that Z_{12}, as given by the latter two equations, may be used with the previously derived formula for the self-radiation impedance of a single rhombic antenna[3] for substitution into Eq. (6), giving the free space radiation impedance of two identical rhombic antennas in cascade.

Perhaps it should be emphasized that although the radiation impedance of a system of rhombic antennas determines the power radiated by the system, and is thereby useful in determining the gain of the system, it does not constitute the driving point impedance of the system. However, it does largely determine the attenuation of the current along the various legs of the system, and as a consequence does enter into the finding of the driving point impedance to a certain extent.

The details of the integrations involved may be found in the technical report[4] from which this paper is taken.

[4] U. S. Naval Postgrad. School Tech. Rpt. No. 7 (Dec., 1952).

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Slot Coupling of Rectangular and Spherical Wave Guides*†

L. B. Felsen and N. Marcuvitz
Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, New York

(Received September 8, 1952)

A dominant-mode wave guide radiating through a slot into a half-space may be regarded from a network viewpoint, with the half-space represented by a number of spherical transmission lines, the wave-guide feed by a single uniform transmission line and the slot by a coupling network. This paper contains (a) an investigation of how many spherical transmission lines are necessary to represent adequately the far field (gain pattern) of the slot antenna, and (b) an approximate theoretical evaluation (by a variational calculation) and experimental check of the equivalent circuit parameters of a slot-coupled junction of a rectangular and spherical wave guide, when the far field can be well represented by the dominant spherical mode. The results are useful in connection with a method of measuring the electromagnetic scattering properties of obstacles located in a half-space illuminated by a wave-guide-fed slot antenna.

I. INTRODUCTION

THE problem to be discussed arises in connection with a method of measuring scattering of electromagnetic waves by an obstacle located in a half-space. A wave-guide-fed slot antenna in an "infinite" ground plane is assumed to illuminate the obstacle. In order to determine the scattered electromagnetic fields within such a configuration, it is necessary that the differential scattering cross section of the obstacle be known. The theoretical evaluation of the scattering cross section for obstacles of arbitrary geometry poses extreme difficulties, especially in the range where the obstacle dimensions are comparable to the wavelength. Alternately, the usual experimental determination of the scattering cross section necessitates a tedious measurement of the electromagnetic fields at all points in the space surrounding the obstacle. A much more desirable procedure is to perform all measurements in the input guide. The subject matter of this paper arose in this connection.

In a microwave network interpretation of the above scattering problem the electromagnetic fields within the half-space are represented rigorously by an infinite number of spherical transmission lines and within the feeding wave guide by an infinite number of uniform transmission lines. The characteristic impedance and propagation constant as well as the variation of mode voltage and current along each transmission line are well defined. The effect of the obstacle can be represented as a lumped equivalent circuit that couples in general each transmission line. The virtue of this network reformulation of the electromagnetic scattering problem is that the equivalent circuit parameters, and hence the scattering cross section, of the obstacle in the half-space can be determined formally by standard microwave measurements in the uniform guide. The physical configuration and its rigorous network equivalent are shown in Fig. 1. In the network picture it has been assumed that, in complete generality, an infinity of modes *propagate* in the half-space; however, as in the uniform guide (assumed to propagate only the lowest mode), a small number of modes usually suffices to specify the far fields. The equivalent circuit for the slot is defined between "far" terminals T_1 in the input line and "far" output terminals $T_2' \cdots T_2^n \cdots$ in each of the half space transmission lines. Terminals $T_2' \cdots T_2^n \cdots$ are joined to the input terminals $T_3' \cdots T_3^n \cdots$ of the obstacle equivalent circuit by lengths of transmission line; across the output terminals $T_4' \cdots T_4^n \cdots$ of the

* The contents of this paper were presented before the National Convention of the Institute of Radio Engineers in March 1951.

† This work was carried out under contract AF 19(122)-3 sponsored by the Air Force Cambridge Research Laboratories.

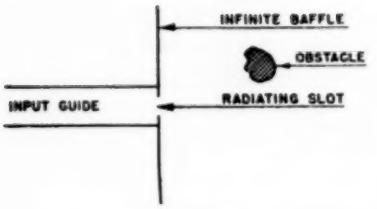


FIG. 1(a). Physical configuration.

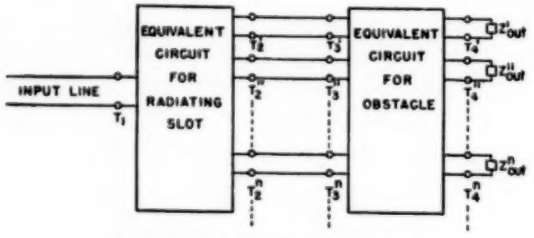


FIG. 1(b). Network equivalent.

obstacle equivalent circuit appear terminating impedances $Z_{out}' \cdots Z_{out}^{n} \cdots$.

If the equivalent circuit coupling the slot to the half-space is known theoretically or otherwise, the equivalent circuit for the obstacle can be found experimentally. Input impedances at T_1 in the input guide are measured for a sufficient number of different terminations Z_{out}, and the resultant input-output relations solved for the desired circuit parameters. The advantage of such a scheme is that measurements are taken only in the input guide where they can be easily performed. Once the obstacle equivalent circuit parameters are known, the "far" scattered fields may be obtained by modal synthesis.

Although the above-described equivalent network procedure affords a formally rigorous approach to the solution of general scattering problems, its practical applicability becomes increasingly difficult when the number of propagating modes which contribute substantially to the far-field configuration in the half-space is large. It is for this reason that one employs a spherical field representation in the half-space which minimizes the number of necessary modes. One notes that a small radiating slot in an obstacle-free half-space excites a far field that behaves approximately like a traveling transverse spherical wave, i.e., a wave which decays in amplitude inversely as the distance from the slot and which has spherical equiphase surfaces centered about the slot. Hence, a field representation in terms of spherical modes adequately describes the far-field behavior in the half-space in terms of only a few spherical transmission lines. As will be discussed below, many suitably excited large slots radiate into the half-space fields that can be approximated for certain applications in terms of a small number of spherical modes, and often in terms of only a single spherical mode.

When a single spherical transmission line description of the half-space is valid, it is of interest to know the equivalent circuit for the slot which couples the input rectangular guide with the output spherical guide (half-space). This paper concerns itself with (1) the conditions under which a single spherical transmission line description is valid, and (2) the theoretical evaluation and experimental verification of the equivalent circuit parameters of a symmetric rectangular slot coupling a rectangular wave guide with a half-space. It should be noted that if a small obstacle is placed in a half-space, the far-zone scattered fields may be represented either as a single spherical wave centered about the obstacle or as an infinity of spherical modes centered about the slot. If the field description in the half-space is carried out in terms of spherical waves centered on the slot, an infinite number of spherical transmission lines will be necessary in the network description even though only one spherical mode is incident on the obstacle. If a single-mode description of the far fields in the half-space is to apply,

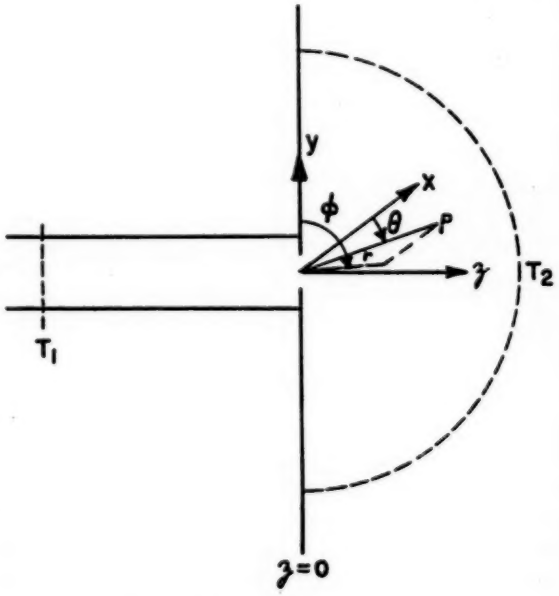


FIG. 2. Orientation of coordinates.

the obstacle must be such that it scatters predominantly the same type of mode as that which is incident upon it, i.e., a spherical wave relative to the slot region. Such obstacles are large structures with a concave spherical curvature relative to the origin. For small obstacles an alternate procedure in terms of spherical modes centered both about the slot and the obstacle is possible.

II. DERIVATION OF THE COUPLING SLOT EQUIVALENT CIRCUIT PARAMETERS

The equivalent circuit for the slot relates the electromagnetic fields at a "far" terminal plane T_1 in the input guide with those at a (hemispheric) "far" terminal plane T_2 in the half-space considered as a spherical wave guide (Fig. 2). The input guide is the familiar (uniform) rectangular wave guide; the output spherical guide is a nonuniform wave guide whose transverse

cross sections are hemispheres centered about the slot center and in which propagation takes place along the radial direction. Since the rectangular guide is assumed to propagate only the dominant H_{10} mode, the transverse vector fields at the terminal plane T_1 far removed from the slot may be represented as

$$\begin{aligned}\mathbf{E}_{t1}(x, y, z_0) &\cong V_1(z_0)\mathbf{e}_1(x, y),\\ \mathbf{H}_{t1}(x, y, z_0) &\cong I_1(z_0)\mathbf{h}_1(x, y).\end{aligned} \tag{2.1a}$$

V_1 and I_1 are the dominant-mode voltage and current, respectively, at terminal plane $T_1(z=z_0)$, and $\mathbf{e}_1(x, y)$ and $\mathbf{h}_1(x, y)$ are the transverse vector mode functions appropriate to the H_{10} mode in rectangular wave guide. In the half-space region an infinity‡ of spherical modes can be propagated; but the transverse vector fields may be approximated at a far terminal plane T_2 by[1]

$$\begin{aligned}r_0\mathbf{E}_{t2}(r_0, \theta, \phi) &\cong V_2(r_0)\mathbf{e}_2(\theta, \phi)+\sum_i^N \mathcal{V}_i(r_0)\mathfrak{e}_i(\theta, \phi),\\ r_0\mathbf{H}_{t2}(r_0, \theta, \phi) &\cong I_2(r_0)\mathbf{h}_2(\theta, \phi)+\sum_i^N \mathcal{I}_i(r_0)\mathfrak{h}_i(\theta, \phi),\end{aligned} \tag{2.1b}$$

where $\mathcal{V}_i(r_0)$ and $\mathcal{I}_i(r_0)$ are the ith spherical-mode voltage and current, respectively, at terminal plane $T_2(r=r_0)$, $\mathfrak{e}_i(\theta, \phi)$ and $\mathfrak{h}_i(\theta, \phi)$ are the transverse vector mode functions appropriate to the ith mode in the spherical guide [Eqs. (2.13)], and r, θ, ϕ are spherical coordinates with the x axis taken as the polar axis. In view of the linear nature of Eqs. (2.1) and of the Maxwell field equations we may write the relationship between the mode voltages and currents at T_1 and T_2 as follows:

$$\begin{aligned}V_1&=\bar{Z}_{11}I_1+\bar{Z}_{12}I_2+\sum_i^N \bar{Z}_{1i}\mathcal{I}_i,\\ V_2&=\bar{Z}_{21}I_1+\bar{Z}_{22}I_2+\sum_i^N \bar{Z}_{2i}\mathcal{I}_i,\\ &\cdots\\ \bar{V}_N&=\bar{Z}_{N1}I_1+\bar{Z}_{N2}I_2+\sum_i^N \bar{Z}_{Ni}\mathcal{I}_i.\end{aligned} \tag{2.2}$$

The coefficients $\bar{Z}_{ij}$ are called the impedance coefficients for the multiterminal equivalent circuit representative of the coupling slot.

■ It is now assumed (see Sec. III) that out of the "infinity" of spherical modes only one (identified by subscript 2) is strongly excited and is reflected appreciably by obstacles§ placed at terminal plane T_2 or beyond. Thus, all the other spherical transmission lines may be considered to be terminated in a match at T_2. The mode voltages and currents on the latter need not be explicitly exhibited since they are related as follows:

$$\mathcal{V}_i = \mathfrak{z}_{0i}\mathcal{I}_i \tag{2.3}$$

where $\mathfrak{z}_{0i}$ is the characteristic impedance for the ith mode transmission line. Thus, Eqs. (2.2) reduce to

$$\begin{aligned}V_1&=Z_{11}I_1+Z_{12}I_2,\\ V_2&=Z_{21}I_1+Z_{22}I_2,\end{aligned} \tag{2.4}$$

where Z_{11}, Z_{22}, and $Z_{12}=Z_{21}$ are now the impedance parameters of the four-terminal equivalent circuit for the coupling slot. The definition of voltage and current is such that power flows into the network from both sides.

To derive explicitly the equivalent circuit parameters for the slot it is necessary to obtain on each side of the slot complete field representations that satisfy the Maxwell equations and possess continuous tangential field components in the slot itself. The desired field representations are determined by replacing the slot discontinuity by an equivalent magnetic current distribution $\mathbf{M}$ flowing on a perfectly conducting surface.[2] Thus, the problem of finding the fields in both region 1 (rectangular wave guide) and region 2 (spherical wave guide) in the presence of the aperture discontinuity may be re-expressed as a simple Kirchhoff-type problem of determining the fields resulting from source distributions in two separate and discontinuity-free geometries. The equivalent source distribution replacing the effect of the slot is related to the slot electric field $\mathbf{E}$ by $\mathbf{M}=\mathbf{E}\times\mathbf{n}$ where $\mathbf{n}$ is a unit vector normal to the slot surface and pointing into the region in question. The slot tangential electric field $\mathbf{E}\times\mathbf{n}$ is not known explicitly

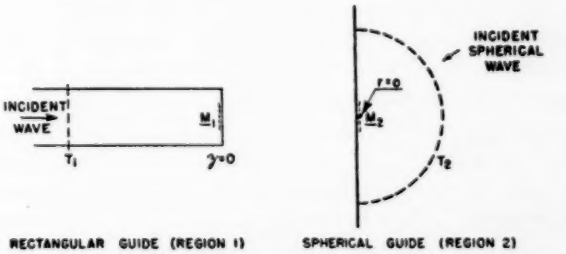


FIG. 3(a). Equivalent geometries.

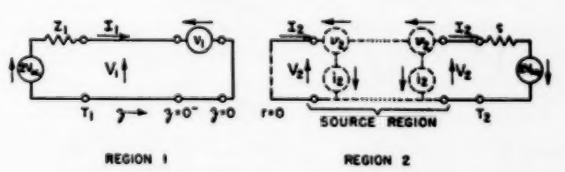


FIG. 3(b). Equivalent network problems.

‡ Actually the number of significant modes radiated by a finite source is finite (see Appendix).

[1] N. Marcuvitz, *Waveguide Handbook* (McGraw-Hill Book Company, Inc., New York, 1951), Sec. 1.8.

§ As stated above only large concave obstacles are treated completely in this manner. Small obstacles require a further treatment.

[2] S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), Secs. 4.5 and 6.13.

but is given implicitly in terms of an integral equation, as will be shown below.

The two separate geometries equivalent to the original configuration in Fig. 2 are shown in Fig. 3(a). The problem in region 1 is that of finding the fields everywhere in a rectangular wave guide terminated in a short circuit at $z=0$ and excited both by a wave incident from $z=-\infty$ and by a magnetic current distribution $\mathbf{M}_1$ at $z=0$. In region 2 the problem is that of determining the fields everywhere in a spherical wave guide (whose transverse cross sections are bounded by the infinite plane) excited by a dominant spherical-mode incident from $r=\infty$ and by a magnetic current distribution $\mathbf{M}_2(=-\mathbf{M}_1)$ on the guide boundary. Since $\mathbf{M}_1=-\mathbf{M}_2$ insures continuity of electric field in the slot, only the requirement of continuity of the tangential magnetic field in the aperture plane remains to be imposed.

Since the aperture plane is transverse relative to region 1, consideration of the transverse fields in region 1 suffices. In terms of a complete orthonormal set of transverse vector mode functions the transverse fields may be represented as

$$\begin{aligned}\mathbf{E}_{t1}(x,y,z)&=V_1(z)\mathbf{e}_1(x,y)+\sum_i{}' V_i(z)\mathbf{e}_i(x,y),\\ \mathbf{H}_{t1}(x,y,z)&=I_1(z)\mathbf{h}_1(x,y)+\sum_i{}' I_i(z)\mathbf{h}_i(x,y).\end{aligned}\tag{2.5}$$

The subscript 1 identifies the dominant H_{10} mode‖ in rectangular guide; the summation is to be taken over all (E and H) modes except the dominant, i.e., over all higher modes. Whenever necessary, the E modes are denoted by single primes and the H modes by double primes. In view of the orthonormality of the vector mode set the mode amplitudes V_i and I_i are given by

$$V_i=\int\int \mathbf{E}_t\cdot\mathbf{e}_i dS_1,\quad I_i=\int\int \mathbf{H}_t\cdot\mathbf{h}_i dS_1,\tag{2.6}$$

where the integration extends over the complete transverse guide cross section S_1. The transverse vector mode functions $\mathbf{e}_i$ and $\mathbf{h}_i$ are derivable for E modes from scalar functions as follows:[3]

$$\mathbf{e}_i'=-\nabla_t\Phi_i,$$

$$\mathbf{h}_i'=\mathbf{z}_0\times\mathbf{e}_i';\quad \nabla_t=\mathbf{x}_0\frac{\partial}{\partial x}+\mathbf{y}_0\frac{\partial}{\partial y},$$

$$\Phi_i=\frac{2}{\pi}\frac{1}{\left(\dfrac{m^2}{a}+\dfrac{n^2}{b}\right)^{\frac{1}{2}}}\sin\frac{m\pi x_1}{a}\sin\frac{n\pi y_1}{b};\tag{2.7a}$$

$$m,\ n=1,\ 2,\ 3\cdots,$$

$$x_1=x+\frac{a}{2},\quad y_1=y+\frac{b}{2};$$

for H modes,

$$\mathbf{h}_i''=-\nabla_t\psi_i;\quad \mathbf{e}_i''=\mathbf{h}_i''\times\mathbf{z}_0,$$

$$\psi_i=\frac{(\epsilon_m\epsilon_n)^{\frac{1}{2}}}{\pi}\frac{1}{\left(\dfrac{m^2}{a}+\dfrac{n^2}{b}\right)^{\frac{1}{2}}}\cos\frac{m\pi x_1}{a}\cos\frac{n\pi y_1}{b},\tag{2.7b}$$

$$m,\ n=0,\ 1,\ 2,\ 3\cdots,\quad \text{mode } m=n=0 \text{ excluded},$$

$$\epsilon_m=1,\ m=0;\quad \epsilon_m=2,\ m\neq 0,$$

where i stands for the double index (m, n), $\mathbf{x}_0$ and $\mathbf{y}_0$ are unit vectors in the x and y directions, a and b are the rectangular guide dimensions parallel to the x and y axes in Fig. 2, respectively. The origin of the coordinate system xyz is located at the guide center.

The mode voltages and currents V_i and I_i satisfy the transmission line equations[4]

$$\begin{aligned}-\frac{dV_i}{dz}&=j\kappa_i Z_i I_i+v_i,\\ -\frac{dI_i}{dz}&=j\kappa_i Y_i V_i+i_i,\end{aligned}\tag{2.8}$$

where for E modes,

$$\kappa_i'=(k^2-k_{ci}'^2)^{\frac{1}{2}},\quad Y_i'=1/Z_i'=\omega\epsilon/\kappa_i',\quad k=\omega(\mu\epsilon)^{\frac{1}{2}},\tag{2.8a}$$

and for H modes,

$$\kappa_i''=(k^2-k_{ci}''^2)^{\frac{1}{2}},\quad Z_i''=1/Y_i''=\omega\mu/\kappa_i'',$$

$$k_{ci}'^2=k_{ci}''^2=\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2.\tag{2.8b}$$

The source terms v_i and i_i in Eqs. (2.8) arise from the presence of the sources in the region. Since, in the present instance, the magnetic source currents $\mathbf{M}_1$ are transverse [Fig. 3(a)] the source current terms i_i are not excited and the source voltages v_i are given by

$$v_i(0)=\int\int_{ap}\mathbf{M}_1(x,y,0)\cdot\mathbf{h}_i(x,y)dS_1;\quad i_i(0)=0,\tag{2.9}$$

where the integration extends over the aperture surface.

The transmission line Eqs. (2.8) for region 1 permit the dominant mode network representation shown in Fig. 3(b). The source term v_1 acts as a voltage generator in the transmission line, and the incident wave from $z=-\infty$ is represented by a voltage generator of strength $2V_{\text{inc}1}$ and internal impedance Z_1. $V_{\text{inc}1}$ is the amplitude of the dominant mode incident wave at T_1, where T_1 is taken to be an integral number of guide wavelengths away from the plane $z=0$. The network applies to the

‖ Throughout this paper the notations E and TM modes, and H and TE modes, respectively, are employed interchangeably.

[3] See reference 1, Sec. 1.2, Eqs. (3), (4), Sec. 2.2, Eqs. (1), (7).

[4] N. Marcuvitz and J. Schwinger, J. Appl. Phys. **22** (1951), pp. 812–813, Eqs. (3.10) and (3.12).

higher modes as well except that $V_{inc_i} \equiv 0$, since only a dominant mode is assumed incident. Thus, for the higher modes, the voltage generator v_i sees a matched load so that the higher mode currents I_i at $z=0^-$ are immediately given by

$$I_i(0^-) = -Y_i v_i(0). \tag{2.10}$$

After substituting for v_i from Eq. (2.9) one writes the transverse magnetic field in Eq. (2.5) at the aperture plane as

$$\mathbf{H}_{t1}(x, y, 0^-) = I_1 \mathbf{h}_1(x, y) - \int\int_{ap} B_g(x, y, 0^-; x', y', 0) \cdot \mathbf{M}_1(x', y', 0) dS', \tag{2.11}$$

where

$$B_g(x, y, 0^-; x', y', 0) = \sum_i{}' Y_i \mathbf{h}_i(x, y) \mathbf{h}_i(x', y') \tag{2.11a}$$

is the dyadic susceptance of guide 1. The integration variables in the aperture have been primed. B_g is an imaginary quantity since for the higher modes the propagation constant κ_i, and therefore Y_i, is imaginary. $I_1(0^-)$ is identical with I_1, the dominant mode current at T_1, since T_1 is an integral number of guide wavelengths away from $z=0$.

The determination of the fields in region 2 is somewhat more complicated since, relative to the spherical guide, the aperture is located on the guide boundary so that it constitutes a longitudinal discontinuity. The tangential magnetic field at the slot is thus not transverse relative to the spherical guide so that an expression for the total magnetic field must be developed. The total electromagnetic fields in the spherical wave guide may be represented in terms of a complete orthonormal set of transverse spherical vector mode functions as follows:[1]

$$\begin{aligned} r\mathbf{E}(r, \theta, \phi) &= V_2(r)\mathbf{e}_2(\theta, \phi) + Z_{s2} I_2(r) \mathbf{e}_{r2}(\theta, \phi) \\ &\quad + \sum_i{}' [\mathcal{V}_i(r)\mathbf{e}_i(\theta, \phi) + \mathfrak{z}_{si} \mathcal{I}_i(r) \mathbf{e}_{ri}(\theta, \phi)], \\ r\mathbf{H}(r, \theta, \phi) &= I_2(r)\mathbf{h}_2(\theta, \phi) + Y_{s2} V_2(r) \mathbf{h}_{r2}(\theta, \phi) \\ &\quad + \sum_i{}' [\mathcal{I}_i(r)\mathbf{h}_i(\theta, \phi) + \mathfrak{y}_{si} \mathcal{V}_i(r) \mathbf{h}_{ri}(\theta, \phi)], \end{aligned} \tag{2.12}$$

$$\mathfrak{z}_{si} = \frac{-jn(n+1)}{r\omega\epsilon}; \quad \mathfrak{y}_{si} = \frac{-jn(n+1)}{r\omega\mu}, \quad i = (m, n).$$

The subscript 2 identifies the "dominant" H_{01} mode (Section III) in the spherical guide (i.e., that mode which is so strongly excited that it alone need be considered at the far terminal plane T_2). The summation extends over all E and H modes except the dominant mode. The script lettering designates higher mode quantities in the spherical guide. (Wherever it is necessary, the E modes will be distinguished by single primes and the H modes by double primes.) The spherical vector mode functions are derivable from scalar functions as follows:[1]

$$\mathbf{e}_i' = -r_t \nabla \Phi_{is}; \quad \mathbf{h}_i' = \mathbf{r}_0 \times \mathbf{e}_i'; \quad \mathbf{e}_{ri}' = \mathbf{r} \frac{\nabla_t \cdot \mathbf{e}_i'}{n(n+1)}; \quad \mathbf{h}_{ri}' \equiv 0, \tag{2.13a}$$

$$\Phi_{is} = \frac{1}{N_i} P_n^m(\cos\theta) \sin m\phi; \quad m = 0, 1, 2 \cdots n; \quad n = 1, 2 \cdots$$

for E modes, and

$$\mathbf{h}_i'' = -r_t \nabla \psi_{is}; \quad \mathbf{e}_i'' = \mathbf{h}_i'' \times \mathbf{r}_0; \quad \mathbf{h}_{ri}'' = \mathbf{r} \frac{\nabla_t \cdot \mathbf{h}_i''}{n(n+1)}; \quad \mathbf{e}_{ri}'' \equiv 0, \tag{2.13b}$$

$$\psi_{is} = \frac{1}{N_i} P_n^m(\cos\theta) \cos m\phi; \quad m = 0, 1, 2 \cdots n; \quad n = 1, 2 \cdots$$

for H modes. The transverse gradient is defined as either

$$_t\nabla = \boldsymbol{\theta}_0 \frac{1}{r} \frac{\partial}{\partial\theta} + \boldsymbol{\phi}_0 \frac{1}{r \sin\theta} \frac{\partial}{\partial\phi},$$

or

$$\nabla_t = \left(\frac{1}{r \sin\theta} \frac{\partial}{\partial\theta} \sin\theta \boldsymbol{\theta}_0\right) + \left(\frac{1}{r \sin\theta} \frac{\partial}{\partial\phi} \boldsymbol{\phi}_0\right) \tag{2.13c}$$

and the normalization constant N_i is given by

$$N_i^2 = \frac{2\pi}{\epsilon_m} \frac{n(n+1)}{(2n+1)} \frac{(n+m)!}{(n-m)!}, \quad \epsilon_m = \begin{cases} 1, & m=0 \\ 2, & m \neq 0 \end{cases}, \tag{2.13d}$$

$\mathbf{r}_0 = \mathbf{r}/r$, $\boldsymbol{\theta}_0$, $\boldsymbol{\phi}_0$ are unit vectors in the r, θ, ϕ directions, respectively, and $P_n^m(\cos\theta)$ is the associated Legendre function of order n and degree m. The scalar mode functions Φ_{is} and ψ_{is} are appropriate to the coordinates shown in Fig. 2, with the polar (x) axis in the infinite plane. They differ from the free-space mode functions in the *Waveguide Handbook* [Sec. 2.8, Eq. (89)] in view of the different boundary conditions imposed by the half-space. The half-space could also have been characterized by a spherical coordinate system whose polar axis is perpendicular to the infinite plane, thus treating the region as a conical wave guide of aperture $\theta_0 = \pi/2$. However, the former choice of coordinates leads to a simplified description of the mode picture in the half-space (see Sec. III).

The spherical mode voltages $\mathcal{V}_i$ and currents $\mathcal{I}_i$ satisfy transmission line equations analogous to those in

Eqs. (2.8):

$$-\frac{d\mathcal{V}_i}{dr}=j\kappa_{is}\mathfrak{z}_i\mathcal{I}_i+\mathfrak{v}_i,$$
$$-\frac{d\mathcal{I}_i}{dr}=j\kappa_{is}\mathcal{Y}_i\mathcal{V}_i+\mathfrak{i}_i, \tag{2.14}$$

where, for E modes,

$$\kappa_{is}'=\kappa_{is}'(r)=\left[k^2-\frac{n(n+1)}{r^2}\right]^{\frac{1}{2}};$$
$$\kappa_{is}'\sim k,\quad kr\gg n;\quad k=\omega(\mu\epsilon)^{\frac{1}{2}}, \tag{2.14a}$$
$$\mathcal{Y}_i'=\mathcal{Y}_i'(r)=1/\mathfrak{z}_i'=\omega\epsilon/\kappa_{is}',$$

and for H modes,

$$\kappa_{is}''=\kappa_{is}',\quad \mathfrak{z}_i''=1/\mathcal{Y}_i''=\omega\mu/\kappa_{is}''. \tag{2.14b}$$

Equations (2.14) apply also to the dominant mode. The dependence of the propagation constant κ_{is} and characteristic impedance $\mathfrak{z}_i$ on the longitudinal variable r indicates that a spherical transmission line is nonuniform. The source terms $\mathfrak{v}_i$ and $\mathfrak{i}_i$ may be related to the magnetic source current distribution $\mathbf{M}_2$ as in the uniform guide case (see Appendix). As in the latter case the transmission line Eqs. (2.14) permit a schematic representation which is shown in Fig. 3(b) for the dominant mode. The calculation of the spherical mode voltages and currents in the aperture in region 2 is more complicated than the corresponding calculation in region 1 since for the latter, the aperture is not confined to a transverse surface. It is shown in the Appendix [Eq. (A23)] that the total magnetic field at any point in the half-space is given by

$$\mathbf{H}_2(\mathbf{r})=-I_2\mathbf{h}_2^{(1)}(\mathbf{r})-\int\int_{ap}[Y_A(\mathbf{r},\mathbf{r}')$$
$$-\eta\mathbf{h}_2^{(1)}(\mathbf{r})\mathbf{h}_2^{(1)}(\mathbf{r}')]\cdot M_2(\mathbf{r}')dS', \tag{2.15}$$

where

$$Y_A(\mathbf{r},\mathbf{r}')=Y_A(\mathbf{r}',\mathbf{r})=\frac{j\omega\epsilon}{2\pi}\left(I+\frac{\nabla\nabla}{k^2}\right)\frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} \tag{2.15a}$$

is the dyadic admittance of the half-space, and

$$r\mathbf{h}_2^{(1)}(\mathbf{r})=\mathfrak{j}_1'(kr)\mathbf{h}_2(\theta,\phi)-j\zeta Y_{s2}\mathfrak{j}_1(kr)\mathbf{h}_{r2}(\theta,\phi),$$
$$\eta=1/\zeta=(\epsilon/\mu)^{\frac{1}{2}},$$
$$\mathfrak{j}_1(kr)=(\pi kr/2)^{\frac{1}{2}}J_{\frac{1}{2}}(kr)=(\sin kr/kr)-\cos kr, \tag{2.15b}$$
$$\mathfrak{j}_1'(kr)=[d/d(kr)]\mathfrak{j}_1(kr).$$

The amplitude $I_2=-I_2(r_0)$ of the dominant mode current refers to the terminal plane T_2 located at a radius $r_0=(4q+1)\lambda/4$, where q is a large integer. $(\mathbf{r})$ stands for (r,θ,ϕ). I is a unit dyadic such that $I\cdot\mathbf{A}=\mathbf{A}\cdot I=\mathbf{A}$.

The dyadic kernel in the integral in Eq. (2.15) consists of the half space dyadic admittance minus another term. The half-space admittance takes into account the entire field scattered by the slot, including that of the dominant spherical mode. The subtracted term represents the real (conductive) part of the dominant mode so that the dominant mode part in the kernel is purely reactive. It is shown in Sec. III that for the slots under consideration the real power carried by the higher modes is very small compared to that carried by the dominant mode. Thus, the real parts of the higher mode terms in the kernel make small contributions to the integral, so that the kernel is approximately reactive (imaginary).

The solutions (2.11) and (2.15) for the magnetic fields in regions 1 and 2 can now be connected by invoking the continuity of the tangential magnetic field in the aperture. Equating the tangential components of $\mathbf{H}$ (or, equivalently, $\mathbf{z}_0\times\mathbf{H}$), we obtain the integral equation for the tangential electric field $\mathbf{E}$ in the slot

$$I_1\mathbf{z}_0\times\mathbf{h}_1(\mathbf{r})+I_2\mathbf{z}_0\times\mathbf{h}_2^{(1)}(\mathbf{r})$$
$$=\mathbf{z}_0\times\int\int_{ap}K(\mathbf{r},\mathbf{r}')\cdot\mathbf{z}_0\times\mathbf{E}(\mathbf{r}')dS', \tag{2.16}$$

$\mathbf{r}$ in aperture, where

$$K(\mathbf{r},\mathbf{r}')=K(\mathbf{r}',\mathbf{r})=Y_A(\mathbf{r},\mathbf{r}')$$
$$+B_g(\mathbf{r},\mathbf{r}')-\eta\mathbf{h}_2^{(1)}(\mathbf{r})\mathbf{h}_2^{(1)}(\mathbf{r}') \tag{2.16a}$$

and where it has been recognized that in view of the continuity of the tangential electric field in the aperture, $\mathbf{M}_1=-\mathbf{E}\times\mathbf{z}_0=-\mathbf{M}_2$. The integral Eq. (2.16) for the total tangential electric aperture field can be reduced to two simpler integral equations for the partial electric fields $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ which are related to the total tangential field by

$$\mathbf{E}(\mathbf{r})=I_1\mathbf{E}^{(1)}(r)+I_2\mathbf{E}^{(2)}(\mathbf{r}). \tag{2.17}$$

Equation (2.17) is justified because of the linear nature of Eq. (2.16). It is apparent that $\mathbf{E}=\mathbf{E}^{(1)}$ when $I_1=1$, $I_2=0$, and $\mathbf{E}=\mathbf{E}^{(2)}$ when $I_2=1$, $I_1=0$. Thus, $\mathbf{E}^{(1)}$ is the aperture field when an unit amplitude current is applied at terminal plane T_1 and an open-circuit at terminal plane T_2; the converse is true for $\mathbf{E}^{(2)}$. Upon substituting Eq. (2.17) into Eq. (2.16) and equating like coefficients of I_1 and I_2 (which are independent of one another), one obtains the following two integral equations:

$$\mathbf{z}_0\times\mathbf{h}_1(\mathbf{r})=\mathbf{z}_0\times\int\int_{ap}K(\mathbf{r},\mathbf{r}')\cdot\mathbf{z}_0\times\mathbf{E}^{(1)}(\mathbf{r}')dS', \tag{2.18a}$$

$\mathbf{r}$ in aperture,

$$\mathbf{z}_0\times\mathbf{h}_2^{(1)}(\mathbf{r})=\mathbf{z}_0\times\int\int_{ap}K(\mathbf{r},\mathbf{r}')\cdot\mathbf{z}_0\times\mathbf{E}^{(2)}(\mathbf{r}')dS'. \tag{2.18b}$$

If Eqs. (2.18) are multiplied in dot product fashion by $\mathbf{E}$

and the result is integrated over the aperture, one obtains the desired network Eqs. (2.4), in view of relation (A17) in the Appendix

$$V_2=\int\int_{ap}\mathbf{z}_0\times\mathbf{E}(\mathbf{r}')\cdot\mathbf{h}_2^{(1)}(\mathbf{r}')dS' \qquad (2.19)$$

and Eqs. (2.17) and (2.6). The impedance parameters are given by

$$\begin{aligned}
Z_{11}&=\int\int_{ap}\mathbf{z}_0\times\mathbf{E}^{(1)}\cdot\mathbf{h}_1 dS\\
&=\int\int_{ap}dS\int\int_{ap}dS'\mathbf{z}_0\times\mathbf{E}^{(1)}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}^{(1)},\\
Z_{22}&=\int\int_{ap}\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot\mathbf{h}_2^{(1)}dS\\
&=\int\int_{ap}dS\int\int_{ap}dS'\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}^{(2)}, \qquad (2.20)\\
Z_{12}&=\int\int_{ap}\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot\mathbf{h}_1 dS\\
&=\int\int_{ap}dS\int\int_{ap}dS'\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}^{(1)}\\
&=\int\int_{ap}dS'\mathbf{h}_2^{(1)}\cdot\mathbf{z}_0\times\mathbf{E}^{(1)}=Z_{21}.
\end{aligned}$$

Z_{11}, Z_{12}, Z_{22} are the desired impedance parameters for the slot coupling the rectangular wave guide with the spherical wave guide. The equivalent circuit is valid between reference plane T_1 in the rectangular guide and T_2 in the spherical guide (see Fig. 4). Reference plane T_1 is located a large integral number of guide wavelengths away from the slot; reference plane T_2 is located at a radius $r_0=(4q+1)\lambda/4$, where q is a large integer.

In order to evaluate the equivalent circuit parameters from Eqs. (2.20) the partial fields $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ must be known, i.e., the integral Eqs. (2.18) must be solved. An exact solution of these integral equations is extremely difficult. Thus, it is desirable to employ a variational formulation for the equivalent circuit parameters which is not sensitive to the exact nature of the fields and hence forms the basis for an approximation procedure.

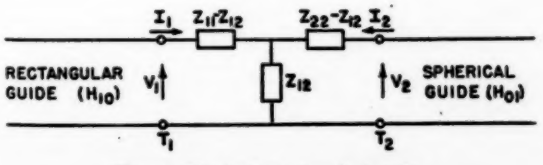


FIG. 4. Equivalent circuit for slot.

Such a formulation is obtained by rewriting Eqs. (2.20) as follows:

$$\begin{aligned}
\frac{1}{Z_{11}}&=\frac{\int\int_{ap}dS\int\int_{ap}dS'\mathbf{z}_0\times\mathbf{E}^{(1)}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}^{(1)}}{\left[\int\int_{ap}dS\,\mathbf{z}_0\times\mathbf{E}^{(1)}\cdot\mathbf{h}_1\right]^2},\\
\frac{1}{Z_{22}}&=\frac{\int\int_{ap}dS\int\int_{ap}dS'\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}^{(2)}}{\left[\int\int_{ap}dS\,\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot\mathbf{h}_2^{(1)}\right]^2} \qquad (2.21)\\
\frac{1}{Z_{12}}&=\frac{\int\int_{ap}dS\int\int_{ap}dS'\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}^{(1)}}{\left[\int\int_{ap}dS\mathbf{z}_0\times\mathbf{E}^{(2)}\cdot\mathbf{h}_1\right]\left[\int\int_{ap}dS\mathbf{z}_0\times\mathbf{E}^{(1)}\cdot\mathbf{h}_2^{(1)}\right]}.
\end{aligned}$$

The variational expressions (2.21) for the impedance parameters are not only independent of the magnitude of both $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ but are independent of first-order variations of the fields about their correct form as well. This latter property is a consequence of the integral equations (2.18) and (2.20).

III. SPHERICAL MODES EXCITED BY THE RECTANGULAR SLOT

In Sec. II the equivalent circuit parameters have been derived for a slot coupling a rectangular wave guide with a half-space, the latter being considered as a spherical wave guide. The derivation is based on the assumption that the half-space can be regarded as a single spherical transmission line; it is implied hereby that a dominant mode is excited so strongly in comparison with all the other modes that it alone is needed to approximately specify the fields at the far terminal plane T_2. In this section it will be shown that this one mode assumption is justified for centered, narrow, rectangular slots, especially if the slots are short. It is also shown that two spherical modes specify the far fields almost completely for longer slots.

In order to determine how many spherical modes are needed to specify the far fields radiated into the half-space by a centered narrow rectangular slot, the spherical mode and closed form representations of the far fields will be compared. The radiated field is given in closed form in Eqs. (2.15) and (2.15a), with a dominant spherical mode incident from infinity. For the present considerations, no incident field need be contemplated; the spherical guide is matched at terminal plane T_2,

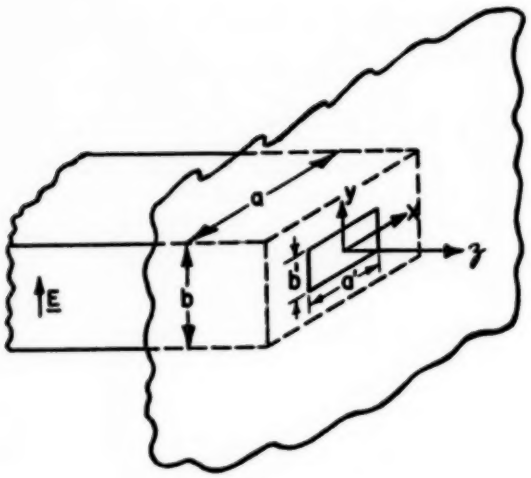


FIG. 5. Slot geometry.

i.e., $-I_2=\eta V_2$. In view of Eqs. (2.15) and (2.19), the field everywhere radiated by the slot is, therefore, given by

$$\mathbf{H}_2(\mathbf{r})=\int\int_{ap} Y_h(\mathbf{r},\mathbf{r}')\cdot\mathbf{z}_0\times\mathbf{E}(\mathbf{r}')dS'. \tag{3.1}$$

In the far region $\mathbf{r}\gg\mathbf{r}'$, $|\mathbf{r}-\mathbf{r}'|\approx r-r'\cos\gamma$, where γ is the angle between the vector $\mathbf{r}'$ in the aperture plane and the vector $\mathbf{r}$ defining the far observation point. Thus, to order $1/r$, the far magnetic field may be written in closed form as

$$\mathbf{H}_2(\mathbf{r})\sim\frac{j\omega\epsilon}{2\pi}\frac{e^{-jkr}}{r}(I-\mathbf{r}_0\mathbf{r}_0)\cdot\int\int_{ap}\mathbf{z}_0\times\mathbf{E}(\mathbf{r}')e^{-jkr'\cos\gamma}dS'. \tag{3.2}$$

The far field is transverse with respect to the r direction since $(I-\mathbf{r}_0\mathbf{r}_0)=(\boldsymbol{\theta}_0\boldsymbol{\theta}_0+\boldsymbol{\phi}_0\boldsymbol{\phi}_0)$ is a transverse unit dyadic. Since the primary purpose of the present section is the comparison of closed form and spherical mode representations of the far field, it is sufficient to specify an appropriate aperture field $\mathbf{z}_0\times\mathbf{E}$. In view of the symmetry of the configuration (see Fig. 5) and the nature of the incident dominant mode field in the rectangular guide, a reasonable assumption is that

$$\mathbf{E}(\mathbf{r}')\approx\mathbf{y}_0E_0\cos(\pi x'/a'), \tag{3.3}$$

where a' and b' are the slot dimensions. The aperture plane is defined in spherical coordinates by $\phi'=0$ for $y'>0$ and $\phi'=\pi$ for $y'<0$. In view of the relations

$$y=r\sin\theta\cos\phi,\ z=r\sin\theta\sin\phi,\quad x=r\cos\theta, \tag{3.4a}$$

one finds that

$$r'\cos\gamma=(\mathbf{r}\cdot\mathbf{r}')/r=y'\sin\theta\cos\phi+x'\cos\theta,\quad\text{for}\quad\phi'=0,\pi. \tag{3.4b}$$

The result of evaluating the simple integrals obtained upon substituting Eqs. (3.3) and (3.4b) into (3.2) is as follows:[5]

$$\mathbf{H}_2(\mathbf{r})\sim\boldsymbol{\theta}_0 f(r)\times\frac{\sin[(kb'/2)\sin\theta\cos\phi]}{(kb'/2)\cos\phi}\frac{\cos[(ka'/2)\cos\theta]}{[1-(ka'/\pi)^2\cos^2\theta]}, \tag{3.5}$$

where

$$f(r)=\frac{j\omega\epsilon a'b'}{\pi^2}\frac{e^{-jkr}}{r}E_0,\quad k=\frac{2\pi}{\lambda}=\omega(\mu\epsilon)^{\frac{1}{2}} \tag{3.5a}$$

and where it has been recognized that $\boldsymbol{\theta}_0\cdot\mathbf{x}_0=-\sin\theta$, $\boldsymbol{\phi}_0\cdot\mathbf{x}_0=0$. The angular function in (3.5) has been normalized so that its maximum magnitude (for $\theta=\phi=\pi/2$) is equal to unity. For narrow slots, (kb') is small, so that (3.5) can be written as

$$\mathbf{H}_2(\mathbf{r})\approx\boldsymbol{\theta}_0 f(r)\sin\theta\frac{\cos[(ka'/2)\cos\theta]}{1-(ka'/\pi)^2\cos^2\theta},\quad (kb')\text{ small}; \tag{3.6a}$$

if a' is small as well,

$$\mathbf{H}_2(r)\approx\boldsymbol{\theta}_0 f(r)\sin\theta;\quad (ka'),\ (kb')\text{ small}. \tag{3.6b}$$

The angular dependence and direction of the far field in (3.6b) is the same as that due to a magnetic dipole located at the origin of coordinates on the infinite plane and directed along the x axis. For longer slots, Eq. (3.6a) indicates that the far field behaves like a dipole field, but modified by a factor whose importance depends on the magnitude of (a'/λ).

Since the far fields are transverse to $0(1/r)$, the spherical mode expansion for the far magnetic field is given by Eq. (2.12) as

$$r\mathbf{H}_2(\mathbf{r})\sim I_2(r)\mathbf{h}_2(\theta,\phi)+\sum_i g_i(r)\mathbf{h}_i(\theta,\phi),\quad r\text{ large}. \tag{3.7}$$

The mode currents computed from the network in Fig. 3(b) are listed in Eqs. (A9) of the Appendix; their far values are readily deduced from these expressions. The magnitudes of the currents relative to one another distinguish the strongly excited modes from the weakly or nonexcited ones. In view of the simple symmetry of the source, it is also possible to select the predominant modes by inspecting the far field structure. One notes the fact that the far field described by Eqs. (3.6) is approximately independent of ϕ and that the modes excited are mainly H modes, since the source is characteristic of a magnetic dipole directed along the x axis at the origin of coordinates (such a dipole excites a transverse electric field). The mode functions for the significant modes are those [see Eq. (2.13b)] for which $m=0$ (symmetry in ϕ) and which are even functions of

[5] See W. R. Smythe, Revs. Modern Phys. **20**, 175 (1948), Eq. (25); result for polar axis perpendicular to the infinite plane.

$\cos\theta$ (symmetry about the plane $\theta=\pi/2$). Since $P_{2l+1}(\cos\theta)$, l integer, is an odd function of $\cos\theta$, its derivative is an even function; therefore, the mode number n must be odd so that the strongly excited modes are H_{0n}, n odd. The lowest mode H_{01}, the magnetic dipole mode, has been designated by the subscript 2 in the text. For the first two lowest H modes the mode functions are, from (2.13b),

$$\mathfrak{h}_{01}''(\theta,\phi)\equiv\mathbf{h}_2(\theta,\phi)=\boldsymbol{\theta}_0\left(\frac{3}{4\pi}\right)^{\frac{1}{2}}\sin\theta, \tag{3.8a}$$

$$\mathfrak{h}_{03}''(\theta,\phi)=\boldsymbol{\theta}_0\frac{1}{4}\left(\frac{7}{6\pi}\right)^{\frac{1}{2}}\sin\theta\,[15\cos^2\theta-3]. \tag{3.8b}$$

The far values of the H mode currents as obtained from Eqs. (A9) in the Appendix are

$$\mathfrak{g}_i''(r)\sim j\eta\mathrm{h}_n^{(2)}(kr)\int\int_{ap}\mathbf{M}_2(\mathbf{r}')\cdot\mathfrak{h}_i''^{(1)}(\mathbf{r}')dS',$$

$$\mathbf{M}_2=\mathbf{E}\times\mathbf{z}_0, \tag{3.9}$$

$$r\mathfrak{h}_i''^{(1)}(\mathbf{r})=\mathrm{j}_n'(kr)\mathfrak{h}_i''(\theta,\phi)-j\zeta Y_{si}\mathrm{j}_n(kr)\mathfrak{h}_{ri}''(\theta,\phi),$$

$$\mathrm{h}_n^{(2)}(kr)\sim j^{n+1}e^{-jkr}. \tag{3.9a}$$

The integrals obtained upon substituting the field (3.3) and the appropriate mode functions into Eqs. (3.9) are readily evaluated if the power series expansions for the spherical Bessel functions j_n are employed. The resulting series converges quite rapidly over the range of slots $b'<a'$. One obtains

$$I_2(r)=\mathfrak{g}_{01}''(r)\sim jE_0\frac{4\omega\epsilon a'b'}{3\pi}\left(\frac{3}{4\pi}\right)^{\frac{1}{2}}e^{-jkr}$$

$$\times\left\{1-\frac{2}{5}\left(\frac{a'}{\lambda}\right)^2\left[\left(\frac{\pi^2}{4}-2\right)+\frac{\pi^2}{6}\left(\frac{b'}{a'}\right)^2\right]\right.$$

$$\left.+0\left[\left(\frac{a'}{\lambda}\right)^4\right]+0\left[\left(\frac{a'b'}{\lambda^2}\right)^2\right]+0\left[\left(\frac{b'}{\lambda}\right)^4\right]\right\}, \tag{3.10a}$$

$$\mathfrak{g}_{03}''(r)\sim-jE_0\frac{8\omega\epsilon a'b'}{35\pi}\left(\frac{7}{24\pi}\right)^{\frac{1}{2}}e^{-jkr}\left(\frac{a'}{\lambda}\right)^2$$

$$\times\left\{4\left(\frac{\pi^2}{4}-2\right)-\frac{\pi^2}{6}\left(\frac{b'}{a'}\right)^2+0\left[\left(\frac{a'}{\lambda}\right)^4\right]\right.$$

$$\left.+0\left[\left(\frac{a'b'}{\lambda^2}\right)^2\right]+0\left[\left(\frac{b'}{\lambda}\right)^4\right]\right\}. \tag{3.10b}$$

Thus, the H_{01} mode contribution to the far field is given by

$$\mathbf{H}_{01}''(\mathbf{r})\sim\frac{I_2(r)\mathbf{h}_2(\theta,\phi)}{r}=\boldsymbol{\theta}_0 f(r)\sin\theta$$

$$\times\left\{1-\frac{2}{5}\left(\frac{a'}{\lambda}\right)^2\left[\left(\frac{\pi^2}{4}-2\right)+\frac{\pi^2}{6}\left(\frac{b'}{a'}\right)^2\right]+\cdots\right\}, \tag{3.11}$$

where $f(r)$ is given in Eq. (3.5a). For small slots Eq. (3.11) is identical with (3.6b). Therefore, the dominant (dipole) mode characterizes almost completely the far field radiated by a small rectangular slot excited by an electric field in the y direction.

The combined contribution to the far field of the H_{01} and H_{03} modes is

$$\mathbf{H}_{01}''+\mathbf{H}_{03}''\sim\frac{1}{r}[I_2\mathbf{h}_2+\mathfrak{g}_{03}''\mathfrak{h}_{03}'']$$

$$=\boldsymbol{\theta}_0\sin\theta f(r)\left\{1-\left(\frac{a'}{\lambda}\right)^2\left[\frac{\pi^2}{12}\left(\frac{b'}{a'}\right)^2\right.\right.$$

$$\left.\left.+\tfrac{1}{2}\cos^2\theta\left(\pi^2-8-\frac{\pi^2}{6}\frac{b'^2}{a'^2}\right)\right]+\cdots\right\}. \tag{3.12}$$

In order to determine how closely Eqs. (3.11) and (3.12) approximate Eq. (3.5), the gain patterns derived from each of these equations have been plotted for a series of narrow slots (Fig. 6). The slot widths are constant and equal to $b'/\lambda=0.095$; the slot lengths are $a'=\lambda/4$, $a'=\lambda/2$, and $a'/\lambda=0.714$. The plots represent the angular variation of the squared field magnitudes in two principal planes: $\theta=\pi/2$ (E plane), and $\phi=\pi/2$ (H plane). It is seen that the quarter-wavelength slot in Fig. 6(a) excites to good approximation only the dominant H_{01} mode. For the half-wavelength and larger slots in Figs. 6(b) and 6(c), the dominant mode approximation becomes progressively poorer. However, even for $a'/\lambda=0.714$ (corresponding to a wide open slot $a'=a$ in a rectangular guide $0.4''\times0.9''$) the discrepancy between the dominant mode and closed form results for the field amplitude in the forward direction is at most 10 percent. Thus, for narrow slots the replacement of the half-space by a single dominant mode transmission line is justified as a first approximation. It is also seen that for all the slot lengths considered, the combined H_{01} and H_{03} modes represent the radiated field very closely. Comparisons were also made for wider slots. For $b'/\lambda=0.19$ the worst discrepancy (for $a'/\lambda=0.714$) between the closed form and dominant mode values for $\mathbf{H}$ in the forward direction is 12 percent; the two-mode approximation agrees to within 3 percent. A marked effect of the greater slot width is a sharpening of the E plane pattern which cannot be accounted for by the azimuthally symmetric H_{01} and H_{03} modes.

An important feature of the two-mode approximation (3.12) is that although the H_{03} mode noticeably in-

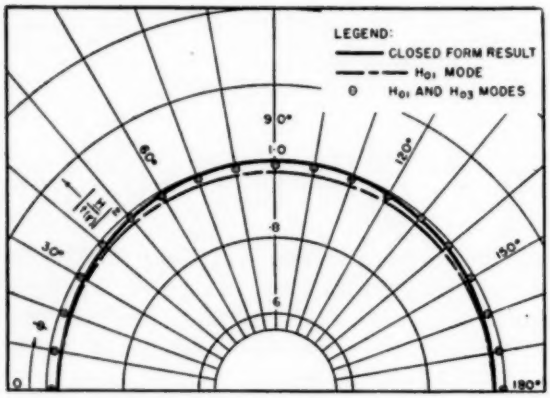


FIG. 6(a-1). E plane; $a'/\lambda=0.250$, $b'/\lambda=0.095$.

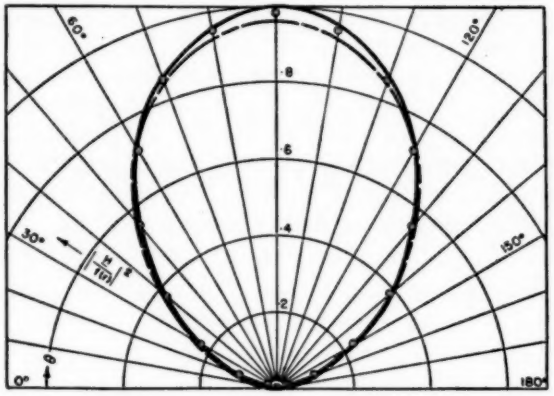


FIG. 6(a-2). H plane; $a'/\lambda=0.250$, $b'/\lambda=0.095$.

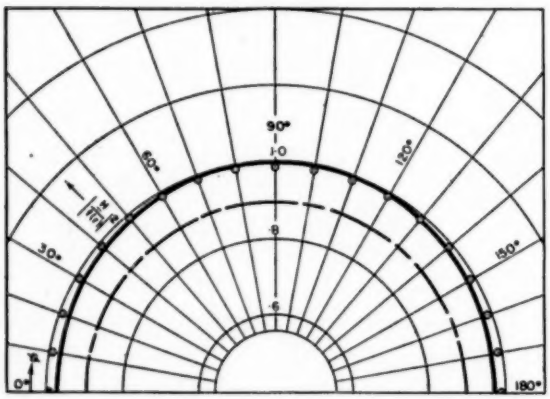


FIG. 6(b-1). E plane; $a'/\lambda=0.500$, $b'/\lambda=0.095$.

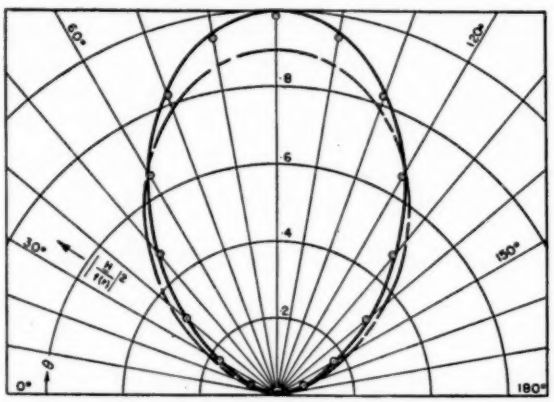


FIG. 6(b-2). H plane; $a'/\lambda=0.500$, $b'/\lambda=0.095$.

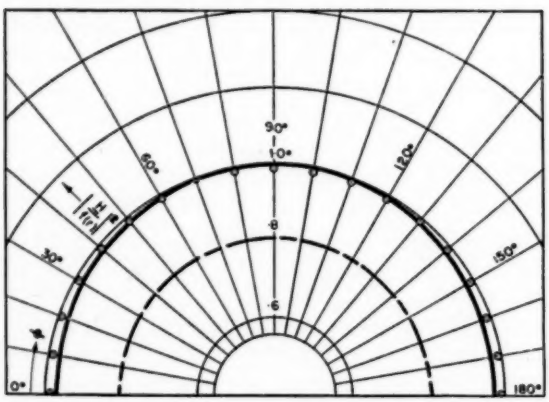


FIG. 6(c-1). E plane; $a'/\lambda=0.714$, $b'/\lambda=0.095$.

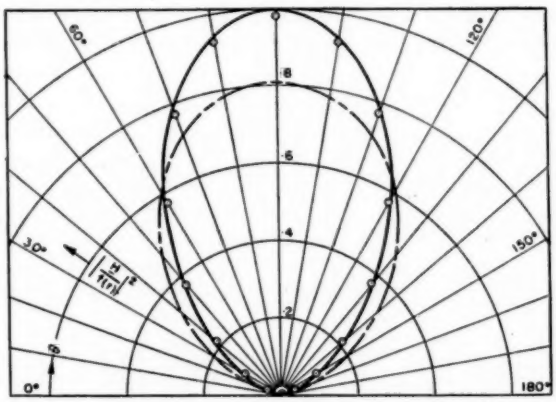


FIG. 6(c-2). H plane; $a'/\lambda=0.714$, $b'/\lambda=0.095$.

FIG. 6. Gain patterns for various slots.

fluences the gain pattern (for long slots) it carries a negligible fraction of the radiated power. The radiated power is given by

$$P=Re\int\int_S \mathbf{E}(\mathbf{r})\times\mathbf{H}^*(\mathbf{r})\cdot\mathbf{r}_0 dS$$

$$=\zeta\int\int_S |\mathbf{H}_t|^2 r^2 d\Omega, \quad d\Omega=\sin\theta d\theta d\phi, \quad (3.13a)$$

where the integration is carried out over a hemisphere with large radius, and the far field relation $\mathbf{E}=\mathbf{E}_t=\zeta\mathbf{H}_t\times\mathbf{r}_0$ has been employed. If the spherical mode expansion (3.7) is substituted into (3.13a) one obtains, in view of the orthonormality of the mode set $\mathfrak{h}_i$ over the angular domain Ω (or, more readily, from the transmission line picture),

$$P=\zeta[|I_2|^2+\sum_i |g_i|^2]$$

$$=\zeta|I_2|^2[1+(|g_{03}''|^2/|I_2|^2)+\cdots]. \quad (3.13b)$$

For the slots considered in Figs. 6a, 6b, 6c, the values of the H_{03}/H_{01} power ratio $|\vartheta_{03}''|^2/|I_2|^2$ are 0.0012, 0.0026, and 0.0124, respectively. The, at first glance, surprising sensitivity of the gain pattern to slight traces of higher mode power is analogous to the corresponding sensitivity of standing wave ratio to slight traces of reflected power. In the former case an "angular" interference between spherical waves obtains, whereas in the latter the interference is between incident and reflected waves.

IV. EVALUATION OF THE EQUIVALENT CIRCUIT PARAMETERS

A. Theoretical Evaluation

The rigorous expressions for the equivalent circuit parameters of the slot coupling the rectangular wave guide to the single-mode spherical wave guide are given in Eqs. (2.21). Their evaluation requires the knowledge of the partial fields $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$. It was pointed out in Sec. II, that $\mathbf{E}^{(1)}$ is the actual slot field set up by an unit dominant mode current at T_1 in the rectangular guide when an open circuit is placed at terminal plane T_2 in the spherical guide. The converse considerations apply to $\mathbf{E}^{(2)}$. In view of the symmetry about the x and y axes of the dominant mode fields in both wave guides, the fields $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ are similar in form. Thus, on use of the approximation that both $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ are proportional to $\mathbf{E}$, Eqs. (2.21) become

$$\frac{1}{Z_{11}} \cong \frac{\int\int_{ap} dS \int\int_{ap} dS' \mathbf{z}_0\times\mathbf{E}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}}{\left[\int\int_{ap} \mathbf{z}_0\times\mathbf{E}\cdot\mathbf{h}_1 dS\right]^2},$$

$$\frac{1}{Z_{12}} \cong \frac{\int\int_{ap} dS \int\int_{ap} dS' \mathbf{z}_0\times\mathbf{E}\cdot K\cdot\mathbf{z}_0\times\mathbf{E}}{\left[\int\int_{ap} \mathbf{z}_0\times\mathbf{E}\cdot\mathbf{h}_1 dS\right]\left[\int\int_{ap} \mathbf{z}_0\times\mathbf{E}\cdot\mathbf{h}_2^{(1)} dS\right]}, \tag{4.1}$$

$$Z_{11}Z_{22}-Z_{12}^2 \cong 0. \tag{4.1a}$$

The validity of the above-mentioned approximation may be tested by examining the inhomogeneous terms of the integral Eqs. (2.18). The vector mode function for the dominant H_{10} mode in rectangular guide is given by [see Eq. (2.7b)]

$$\mathbf{h}_1(x, y) = \mathbf{h}_{01}''(x, y) = \mathbf{x}_0\left(\frac{2}{ab}\right)^{\frac{1}{2}} \cos\frac{\pi x}{a}, \tag{4.2}$$

while that for the dominant spherical mode is given in Eq. (2.15b). Evaluated in a small slot, these mode functions are

$$\mathbf{h}_1 \approx \mathbf{x}_0\left(\frac{2}{ab}\right)^{\frac{1}{2}},$$

$$\mathbf{h}_2^{(1)} \approx \tfrac{2}{3}k\left(\frac{3}{4\pi}\right)^{\frac{1}{2}}(\boldsymbol{\theta}_0\sin\theta - \mathbf{r}_0\cos\theta) = -\tfrac{2}{3}k\left(\frac{3}{4\pi}\right)^{\frac{1}{2}}\mathbf{x}_0, \tag{4.3}$$

i.e., $\mathbf{h}_1 \propto \mathbf{h}_2^{(1)}$, approximately. Thus, the field approximation is justified to first order for small slots.

Equations (4.1) are employed most advantageously in connection not with the circuit shown in Fig. 4, but rather with the alternative circuit in Fig. 7(a) which reduces to that in Fig. 7(b) in view of the approximate relation (4.1a). The parameters of the approximate network in Fig. 7(b) are given by [see (2.16a)]

$$\frac{1}{Z_{11}} = Y_{11} = Y_1 - \eta/n^2, \tag{4.4a}$$

$$Y_1 = \frac{\int\int_{ap} dS \int\int_{ap} dS' \mathbf{z}_0\times\mathbf{E}\cdot[Y_A+B_g]\cdot\mathbf{z}_0\times\mathbf{E}}{\left[\int\int_{ap} dS\ \mathbf{z}_0\times\mathbf{E}\cdot\mathbf{h}_1\right]^2}, \tag{4.4b}$$

$$n = \frac{\int\int_{ap} dS\ \mathbf{z}_0\times\mathbf{E}\cdot\mathbf{h}_1}{\int\int_{ap} dS\ \mathbf{z}_0\times\mathbf{E}\cdot\mathbf{h}_2^{(1)}}. \tag{4.4c}$$

The admittance Y_1 is the (two-terminal) input admittance of the slot radiating into an unterminated half-space (matched spherical transmission line), and its variational evaluation from (4.4b) is known.[6] The term η/n^2 represents the radiation conductance for the dominant spherical mode as seen from the rectangular guide.

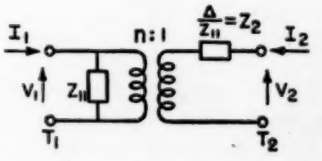


FIG. 7(a). Alternative representation for slot.

$$\Delta = Z_{11}Z_{22} - Z_{12}^2, \quad n = Z_{11}/Z_{12}$$

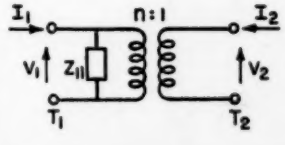


FIG. 7(b). Approximate network.

[6] A. A. Oliner *et al.*, "Equivalent Circuit for Slots in Rectangular Waveguide," Microwave Research Institute, Polytechnic Institute of Brooklyn, August, 1951, Chap. III, Sec. D-4.

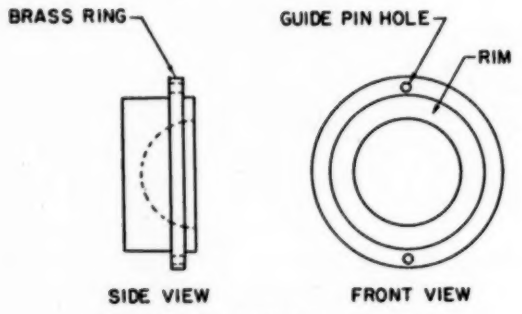


FIG. 8(a). Hemispheric termination.

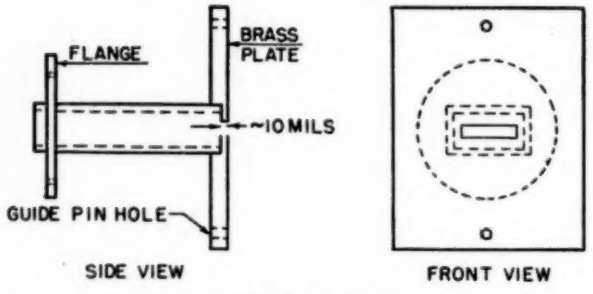


FIG. 8(b). Slot structure.

As pointed out in Sec. III the dominant spherical mode carries by far the major portion of the radiated power, i.e., it makes up the major part of the radiation conductance. Therefore, the shunt admittance Y_{11} should be practically lossless and equal to the radiation susceptance of the slot. This is confirmed by the calculations.

For the approximate evaluation of the turns ratio n, the trial field in Eq. (3.3) will be employed. The numerator integral in Eq. (4.4c) is found to be

$$\iint_{ap} dS' \mathbf{z}_1 \times \mathbf{E} \cdot \mathbf{h}_1 = -E_0 \left(\frac{2}{ab}\right)^{\frac{1}{2}} \int_{-b'/2}^{b'/2} dy' \int_{-a'/2}^{a'/2} dx' \cos\frac{\pi x'}{a'} \cos\frac{\pi x}{a}$$

$$= -E_0 \left(\frac{2}{ab}\right)^{\frac{1}{2}} \frac{2a'b'}{\pi} \frac{1}{1-\alpha^2} \cos\frac{\pi\alpha}{2}, \quad \alpha = \frac{a'}{a}. \tag{4.5}$$

The integral in the denominator of Eq. (4.4c) has already been evaluated in connection with Eq. (3.10a). Thus, the turns ratio n is given by

$$n = -\frac{\lambda}{2}\left(\frac{6}{\pi ab}\right)^{\frac{1}{2}} \times \frac{\cos(\pi\alpha/2)/(1-\alpha^2)}{1-\frac{2}{5}\left(\frac{a'}{\lambda}\right)^2\left[\left(\frac{\pi^2}{4}-2\right)+\frac{\pi^2}{6}\left(\frac{b'}{a'}\right)^2\right]+\cdots} \tag{4.6}$$

n is negative, i.e., V_1 and V_2 are 180° out of phase. It is readily checked that this represents the physical picture correctly. Namely, if V_1 is positive the dominant mode electric field in the rectangular guide $\mathbf{E}_1 = V_1\mathbf{e}_1$ is directed along the $(-y)$ axis. The corresponding dominant mode electric field in the spherical guide $\mathbf{rE}_2 = V_2\mathbf{e}_2 = -|V_2|\mathbf{e}_2$ points in the $+\boldsymbol{\phi}_0$ direction as it should.

B. Experimental Verification

To check the accuracy of the equivalent circuit representation shown in Fig. 7(b) and the corresponding parameter values given by Eqs. (4.4), the four-terminal network between the input plane T_1 and output plane T_2 in Fig. 2 was determined experimentally. Measurements were performed on a series of slots in a rectangular wave guide of inside dimensions $a = 0.900$ in., $b = 0.400$ in. at a free-space wavelength $\lambda = 1.26$ in. $= 3.2$ cm. Four input measurements corresponding to four different output terminations were taken for each slot. The known output terminations were a match (unterminated half-space), and three different lengths of short-circuited spherical wave guide, the latter being realized by three metallic hemispheres of different radius. The four-terminal network as obtained from an analysis of the input-output relations corresponding to the three hemispheric terminations was expressed in the form shown in Fig. 7(a). The network parameters then were checked against a fourth measurement corresponding to the freely radiating slot.

The three hemispheric terminations, shown in Fig. 8(a), were made by boring a hemispheric cavity into solid brass stock on a lathe. The rims of the cups were insulated with thin mica in order to eliminate the possibility of nonuniform contact between the rim and the aperture plane. The aperture structure itself was made especially rigid by first milling the slot into a thick brass plate, soldering the wave guide to it, and then turning the face of the plate down on the lathe until the aperture thickness was about 10 mils. [A diagram of the slot structure is given in Fig. 8(b)]. In this manner, possibility for uneven contact between the insulated cup rims and the slot plane was minimized. The cups were centered with respect to the slot by the use of guide pins and were held to the plane with C-clamps. For the measurement of the freely radiating slot the slot structure was fitted into a hole of the same dimensions in a 4 ft × 6 ft × $\frac{1}{8}$ in. copper plane.

The relationship between the dominant mode voltage reflection coefficient at the origin ($r = 0$) and any other point $r = r_0$ in the spherical guide is given by[1]

$$\Gamma_1(0) = -\Gamma_1(r_0)e^{-j2\tilde{\eta}_1(kr_0)}, \tag{4.7}$$

where

$$\tilde{\eta}_1(kr_0) = \tan^{-1}\left[\frac{n_1(kr_0)}{j_1(kr_0)}\right]; \quad \tilde{\eta}_1(kr) \approx kr - \pi, \quad kr \gg 1. \tag{4.7a}$$

(The subscript 1 refers to the $n = 1$ mode.) If r_0 is taken to be the radius of one of the hemispheric cups, then

$\Gamma_1(r_0)=-1$ so that

$$\Gamma_1(0)=e^{-j2\tilde{\eta}_1(kr_0)}. \tag{4.8}$$

$\Gamma_1(0)$ is particularly simple if $\tilde{\eta}_1$ is chosen to be $-\pi/4$, 0, $\pi/4$. The corresponding values for the reflection coefficient at the origin are

$$\text{(a) } \Gamma_1(0)=+j;\quad \text{(b) } \Gamma_1(0)=+1;\quad \text{(c) } \Gamma_1(0)=-j. \tag{4.9}$$

For the wavelength $\lambda=2\pi/k=1.26$ in., the radii r_0 yielding the values in (4.9) are found to be

$$\text{(a) } r_0=1.067 \text{ in.};\quad \text{(b) } r_0=1.228 \text{ in.};\quad \text{(c) } r_0=1.390 \text{ in.} \tag{4.10}$$

The values in Eq. (4.10) are, of course, not unique; they were chosen for experimental convenience. It is readily seen from Eqs. (4.7) that the dominant mode voltage reflection coefficient at $r=0$ is identical with that at T_2, i.e., at $r=(4q+1)\lambda/4$, $q=$ large integer. For large radii a spherical transmission line behaves like a uniform line; therefore, the input impedance and input reflection coefficient at terminal plane T_2 are related as follows:

$$\frac{Z_{T_2}}{\zeta}=\frac{1+\Gamma_1(0)}{1-\Gamma_1(0)}. \tag{4.11}$$

Thus, one obtains for the output impedances corresponding to the three hemispheric terminations in Eq. (4.10)

$$\text{(a) } \frac{Z_{T_2}}{\zeta}=+j;\quad \text{(b) } \frac{Z_{T_2}}{\zeta}=\infty;\quad \text{(c) } \frac{Z_{T_2}}{\zeta}=-j. \tag{4.12}$$

The corresponding input impedances in the rectangular guide at terminal plane T_1 are

$$\frac{Z_{in_{a,b,c}}}{Z_1}=-j\tan\varphi_{a,b,c},\quad \varphi=\frac{2\pi d}{\lambda_g}, \tag{4.13}$$

where λ_g is the guide wavelength, Z_1 is the guide characteristic impedance, and d is the measured distance of the voltage minimum from terminal plane T_1. The network parameters Z_{11}, Z_{12}, Z_{22} are obtained from the solution of the three equations:

$$Z_{in_{a,b,c}}=Z_{11}-\frac{Z_{12}^2}{Z_{22}+Z_{T_{2a,b,c}}}. \tag{4.14}$$

The equivalent circuit so obtained is checked against a fourth measurement which yields the input impedance for the slot radiating into the unterminated half-space. These results, together with the theoretical parameter values, are listed in Table I (see Fig. 7 for notation employed.) Although expressions for the radiation admittance Y_1 in Eq. (4.4b) are available,[6] the values listed in Table I are experimental. This was done because the purpose of the spherical cup measurement was not to check the agreement of Eq. (4.4b) with measured results but rather to test the validity of the theoretically deduced approximate equivalent circuit in Fig. 7(b) whose parameters are given by Eqs. (4.4a) and (4.4c).

The theoretical value of Z_2' in Table I vanishes under the assumptions in Sec. IV for a zero thickness slot. The theoretical value of Y_{11}' is listed as purely reactive. Actually slight real parts of the order of 0.02 were obtained but discarded, since that order of magnitude corresponds to the tolerances put on the measured values of Y_1. The finite measured value for Z_2' as well as the discrepancy between theoretical and experimental Y_{11}' is at least partly due to the physical thickness of the slot. There also exist differences between the directly measured input admittances for the unterminated half-space and those obtained by terminating the equivalent circuit in a spherical line match. (Where the directly measured input admittances in Table I do not have real parts, the V.S.W.R. was too high to be measured accurately and was arbitrarily taken to be infinite.) The differences between the two results can be attributed to a great extent to the tolerances on the equivalent circuit parameters. The tolerances resulting from error in the input data alone were obtained by computing the sensitivity of each parameter to small deviations in location of the voltage minimum in the rectangular guide and multiplying by the expected possible error. These tolerances were found to be of the order of ± 0.01 on Y_{11}', of the order of ± 0.02 on n, and of the order of $\pm(0.25Z_2')$ on Z_2'. In addition to errors in the input data, there is the possibility of error in the output quantities such as to make the output impedances different from those listed in Eq. (4.12). Such deviations can be brought about by a slightly different radius or nonspherical shape for the spherical cups or by noting the effect of the thin insulating mica strip on the cup rims. Computations to take into account the former sources of output error could be made, but it was felt

TABLE I. Comparison between measured and theoretical values. $Y_{11}'=Y_{11}/Y_1$, $Z_2'=Z_2/Z_1$, $Y_1=1/Z_1=\eta\lambda/\lambda_g$.

Slot size		3-cup measurement			Theoretical		Input admittance (unterminated half-space)	
b'/b	a'/a	Y_{11}'	n	Z_2'	Y_{11}'	n	Measured	Computed from equiv. circuit
0.4	0.4	$-j6.41$	1.35	$+j0.012$	$-j5.75$	1.44	$?-j6.58$	$0.75-j6.43$
	0.5	$-j2.59$	1.35	$+j0.047$	$-j2.30$	1.42	$?-j2.80$	$0.76-j2.64$
	0.6	$-j0.68$	1.36	$+j0.047$	$-j0.61$	1.40	$0.77-j0.67$	$0.75-j0.72$
	0.7	$+j0.196$	1.34	$+j0.037$	$+j0.18$	1.35	$0.81+j0.18$	$0.77+j0.16$

that the nonprecision nature (three-point method) of the measurements did not warrant it. At any rate, the agreement between the quantities compared in Table I is considered reasonable.

APPENDIX

Fields Produced by Arbitrary Sources in a Spherical Wave Guide

The formalism developed for the fields produced by arbitrary sources in a uniform wave guide (see reference 4) may be applied as well to the fields produced by arbitrary sources in a nonuniform spherical wave guide.[7,8] In the latter case the Maxwell field equations are separated with respect to the longitudinal variable r to yield a pair of equations involving only the transverse (to r) components of the electromagnetic fields and their sources. If the transverse part of the modal representation in Eqs. (2.12), with the orthonormal set of modes defined in Eqs. (2.13), is substituted into the transverse field equations, one finds, after some manipulation, that the spherical mode voltages and currents satisfy the inhomogeneous transmission line Eqs. (2.14), where the source terms v_i and i_i are given by[7,8]

$$v_i(r)=\int\int_\Omega r\mathbf{M}(r,\theta,\phi)\cdot\mathbf{h}_i(\theta,\phi)d\Omega - Z_{ai}\int\int_\Omega r\mathbf{J}(r,\theta,\phi)\cdot\mathbf{e}_{ri}(\theta,\phi)d\Omega,$$
$$i_i(r)=\int\int_\Omega r\mathbf{J}(r,\theta,\phi)\cdot\mathbf{e}_i(\theta,\phi)d\Omega - Y_{ai}\int\int_\Omega r\mathbf{M}(r,\theta,\phi)\cdot\mathbf{h}_{ri}(\theta,\phi)d\Omega, \quad \text{(A1)}$$

$d\Omega=\sin\theta d\theta d\phi$ is an element of area in the transverse domain Ω. Since confusion with uniform guide quantities does not arise in this Appendix, ordinary symbols are employed throughout for the spherical transmission line quantities.

For the case of a magnetic source located at $r=r'$ on the infinite plane defined by $\phi=0,\pi$ one has

$$\mathbf{M}(\mathbf{r})=\mathbf{M}_2(r',\theta,\phi)\frac{\delta(r-r')\delta(\phi-\phi')}{r\sin\theta},\quad \phi'=0,\pi,$$
$$\mathbf{J}(\mathbf{r})\equiv 0. \quad \text{(A2)}$$

Thus, the strength of the equivalent voltage and current generators v_i and i_i resulting from a magnetic source at $r=r'$ on the infinite plane is given by

$$v_i(r')=\int_{\theta'}\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i(\theta',\phi')d\theta',\quad \phi'=0,\pi, \quad \text{(A3)}$$
$$i_i(r')=-Y_{ai}\int_{\theta'}\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_{ri}(\theta',\phi')d\theta',$$

where $\mathbf{r}'=(r',\theta',\phi')$ and the integration is extended over all source points at radius r'.

The equivalent network representing the transmission line equations and containing the generators v_i and i_i is shown in Fig. 3(b). In order to determine the transmission line voltage V_i and current I_i at every point r in the region resulting from a magnetic source at r', the behavior of voltage and current along an homogeneous spherical transmission line must be known. The standing and traveling wave solution of the homogeneous spherical transmission line equations (with $v_i=0=i_i$) is given in terms of the spherical Bessel functions by

$$\left.\begin{matrix}I_i'(r)\\ V_i''(r)\end{matrix}\right\}=A_i\,\mathrm{j}_n(kr)+B_i\,\mathrm{n}_n(kr)=C_i\,\mathrm{h}_n^{(1)}(kr)+D_i\,\mathrm{h}_n^{(2)}(kr), \quad \text{(A4a)}$$

$$\left.\begin{matrix}-j\eta V_i'(r)\\ -j\zeta I_i''(r)\end{matrix}\right\}=A_i\,\mathrm{j}_n'(kr)+B_i\,\mathrm{n}_n'(kr)=C_i\,\mathrm{h}_n'^{(1)}(kr)+D_i\,\mathrm{h}_n'^{(2)}(kr), \quad \text{(A4b)}$$
$$z_n(kr)=(\pi kr/2)^{\frac{1}{2}}Z_{n+\frac{1}{2}}(kr),$$

where Z stands for any of the solutions of Bessel's equation. In contrast to a uniform line, the variation of the voltage along a spherical transmission line differs from that of the current for the same mode. It is of interest to point out the analogous role of the propagation constant in uniform and spherical transmission lines. We consider the function $\mathrm{h}_n^{(2)}(kr)$ (or its derivative) which characterizes a spherical traveling wave excited by sources at $r'<r$. For $kr\gg n$, $\mathrm{h}_n^{(2)}$ behaves like a traveling wave of unit amplitude; for $kr\ll n$, $\mathrm{h}_n^{(2)}\propto\mathrm{n}_n\propto\exp[-n\log(2kr/2n-1)]$ decreases as r increases until the transition region $kr\approx n$ is reached. Thus, for $n\ll kr$, the waves "propagate" undamped (except for the $1/r$ variation which is implicit throughout); for $n\gg kr$, the waves are attenuated. The "propagating" and "nonpropagating" ranges correspond to real and imaginary values, respectively, of the propagation constant κ_i in direct analogy to the situation in uniform guides.

The relative input admittance (Y_{i+}/η) [impedance (Z_{i+}/ζ)], as seen from a point r, of a transmission line unterminated (matched) in the $+r$ direction is given

[7] N. Marcuvitz, "Field representations in spherically stratified regions," Comm. on Pure and Applied Mathematics, Vol. IV, August, 1951, Sec. 2b, pp. 277–284.

[8] L. Felsen, "Spherical Transmission Line Theory," Report R-253-31, PIB-194, Microwave Research Institute, Polytechnic Institute of Brooklyn, January, 1952, pp. 43–48.

by

$$\left.\begin{aligned}\frac{I_i'(r)}{\eta V_i'(r)}&=\frac{Y_{i+}'(r)}{\eta}\\ \frac{V_i''(r)}{\zeta I_i''(r)}&=\frac{Z_{i+}''(r)}{\zeta}\end{aligned}\right\}=-j\frac{\mathrm{h}_n^{(2)}(kr)}{\mathrm{h}_n'^{(2)}(kr)}. \tag{A5}$$

Since a spherical transmission line ends at $r=0$, the resulting reflection at $r=0$ demands that the input admittance Y_{i0} (impedance Z_{i0}) looking in the $-r$ direction from a point r be expressed in terms of standing waves (assuming noninfinite values of V and I at $r=0$) as

$$\frac{Y_{i0}'(r)}{\eta}=\frac{Z_{i0}''(r)}{\zeta}=+j\frac{\mathrm{j}_n(kr)}{\mathrm{j}_n'(kr)}, \tag{A6}$$

$$Y_{i0}'(0)=Z_{i0}''(0)=0. \tag{A6a}$$

Thus, the origin behaves like an open circuit for E modes and like a short circuit for H modes.

In view of Eqs. (A4)–(A6), the E-mode currents at any point in the spherical transmission line represented in Fig. 3(b) resulting from the generators v_i and i_i at $r=r'$ are found to be[9]

$$\begin{aligned}\bar{I}_i'(r,r')&=\mathrm{j}_n(kr)[ji_i'(r')\mathrm{h}_n'^{(2)}(kr')\\ &\qquad-\eta v_i'(r')\mathrm{h}_n^{(2)}(kr')], \quad r<r',\\ \bar{I}_i'(r,r')&=\mathrm{h}_n^{(2)}(kr)[ji_i'(r')\mathrm{j}_n'(kr')\\ &\qquad-\eta v_i'(r')\mathrm{j}_n(kr')], \quad r>r'.\end{aligned} \tag{A7a}$$

Since $\bar{V}_i$ and $\bar{I}_i$ satisfy homogeneous transmission line equations for $r\neq r'$, the voltages are given by

$$\bar{V}_i'(r,r')=\frac{j\zeta}{k}\frac{d}{dr}\bar{I}_i'(r,r'). \tag{A7b}$$

For the H modes the pertinent equations are derivable from Eqs. (A7) if the following duality replacements are made

$$\bar{V}_i'\to\bar{I}_i''; \quad \bar{I}_i'\to\bar{V}_i''; \quad v_i'\to i_i'', \quad i_i'\to v_i'', \quad \zeta\to\eta, \quad \eta\to\zeta. \tag{A8}$$

FIG. 9. Domains of integration.

[9] See reference 8, Eqs. (119).

The total scattered mode voltages and currents V_i and I_i are obtained by integrating $\bar{V}_i$ and $\bar{I}_i$ over the source region r'. If the expressions (A3) for the source terms are employed, the result is found to be

$$\begin{aligned}V_i'(r)&=-j\mathrm{h}_n'^{(2)}(kr)\int\int_A dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i'^{(1)}(\mathbf{r}')\\ &\qquad-j\mathrm{j}_n'(kr)\int\int_B dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i'^{(+)}(\mathbf{r}'),\\ I_i'(r)&=-\eta\mathrm{h}_n^{(2)}(kr)\int\int_A dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i'^{(1)}(\mathbf{r}')\\ &\qquad-\eta\mathrm{j}_n(kr)\int\int_B dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i'^{(+)}(\mathbf{r}'),\\ V_i''(r)&=j\mathrm{h}_n^{(2)}(kr)\int\int_A dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i''^{(1)}(\mathbf{r}')\\ &\qquad+j\mathrm{j}_n(kr)\int\int_B dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i''^{(+)}(\mathbf{r}'),\\ I_i''(r)&=-\eta\mathrm{h}_n'^{(2)}(kr)\int\int_A dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i''^{(1)}(\mathbf{r}')\\ &\qquad-\eta\mathrm{j}_n'(kr)\int\int_B dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_i''^{(+)}(\mathbf{r}'),\end{aligned} \tag{A9}$$

where

$$\begin{aligned}r\mathbf{h}_i'^{(-)}_{(+)}(\mathbf{r})&=\mathrm{h}_n^{(1)}_{(2)}(kr)\mathbf{h}_i'(\theta,\phi),\\ r\mathbf{h}_i''^{(-)}_{(+)}(\mathbf{r})&=\mathrm{h}_n'^{(1)}_{(2)}(kr)\mathbf{h}_i''(\theta,\phi)\\ &\qquad-j\zeta Y_{si}''\,\mathrm{h}_n^{(1)}_{(2)}(kr)\mathbf{h}_{ri}(\theta,\phi),\end{aligned} \tag{A9a}$$

and

$$\mathbf{h}_i^{(1)}=\tfrac{1}{2}[\mathbf{h}_i^{(-)}+\mathbf{h}_i^{(+)}]; \quad \mathbf{h}_i^{(2)}=\frac{1}{2j}[\mathbf{h}_i^{(-)}-\mathbf{h}_i^{(+)}].$$

$dS'=r'dr'd\theta'$ is the element of integration in the aperture. The domains of integration A and B are shown in Fig. 9. The magnetic field at any point $\mathbf{r}$ in the half-space, resulting from the magnetic sources in the aperture region, is obtained by substituting Eqs. (A9) into Eqs. (2.12), whence

$$\mathbf{H}(\mathbf{r})=-\int\int_{ap}dS'\mathbf{M}_2(\mathbf{r}')\cdot\sum_i\eta\mathbf{h}_i^{(+)}(\mathbf{r}_>)\mathbf{h}_i^{(1)}(\mathbf{r}_<), \tag{A10}$$

where $\mathbf{r}_>$ denotes the greater and $\mathbf{r}_<$ the lesser of the two coordinates $\mathbf{r}$ and $\mathbf{r}'$. The sum in (A10) represents the spherical mode expansion of the half-space dyadic Green's function. This is readily recognized if one puts

$$\mathbf{M}_2(\mathbf{r}')=\mathbf{1}\,\delta(\mathbf{r}-\mathbf{r}'), \tag{A11}$$

appropriate to a dipole source of unit strength directed along the unit vector $\mathbf{l}$ in the infinite plane and employs the definition of the half-space Green's function Y_h:

$$\mathbf{H}(\mathbf{r})=-Y_h(\mathbf{r},\mathbf{r}')\cdot\mathbf{l}. \tag{A12}$$

It may be recognized from the spherical mode expansion of the half-space dyadic admittance that the number of modes which contribute materially to the far field excited by a given source depends on the dimensions of the source. For $\mathbf{r}\gg\mathbf{r}'$, the $\mathbf{h}_i^{(1)}$ functions characterize the physical extent of the source region through their dependence on $\mathrm{j}_n(kr')$ [or $\mathrm{j}_n'(kr')$]. Investigation of the behavior of $\mathrm{j}_n(kr')$ shows that modes for which $n\ll kr'$ may contribute substantially to the far field, while for $n\gg kr'$ the contributions are insignificant.

So far, only the field scattered by the sources in the slot has been treated. If a dominant mode field of amplitude V_{inc} is incident from infinity as well, it contributes a standing wave whose voltage and current may be written as [see Eqs. (A4) or the equivalent network in Fig. 3(b)]

$$V_u(r)=2C_2\,\mathrm{j}_1(kr), \tag{A13a}$$

where, in view of Eqs. (A4),

$$C_2=C_{01}''=V_{\mathrm{inc}}=(A_2-jB_2)/2.$$

Also

$$I_u(r)=j(\eta/k)(d/dr)V_u(r). \tag{A13b}$$

The subscript u denotes the unperturbed solution (no aperture sources). These unperturbed components must be added to the scattered components in (A9) for the dominant mode to yield the total dominant mode voltage and current. The contribution of the incident wave to the magnetic field in the half-space is given by [note Eqs. (2.12) and (A9a)]

$$\mathbf{H}_u(\mathbf{r})=j\eta 2C_2\mathbf{h}_2^{(1)}(\mathbf{r})=j\eta A_2\mathbf{h}_2^{(1)}+\eta B_2\mathbf{h}_2^{(1)}. \tag{A14}$$

Upon rewriting Eqs. (A4) in asymptotic (large kr) form for the dominant mode, one readily links the coefficients A_2 and B_2 with the total dominant mode voltage and current at the far terminal plane T_2, namely,

$$\begin{aligned} V_2(r)&\sim -A_2\cos kr-B_2\sin kr,\\ -j\zeta I_2(r)&\sim A_2\sin kr-B_2\cos kr. \end{aligned} \tag{A15}$$

Thus, for $kr=kr_0=(4q+1)\pi/2$, q large integer,

$$\begin{aligned} V_2&=-B_2=V_2(r_0),\\ -I_2&=j\eta A_2=I_2(r_0). \end{aligned} \tag{A16}$$

Since $\mathrm{j}_1(kr)\sim-\cos kr=0$ at terminal plane T_2, the unperturbed dominant mode voltage in (A13a) does not contribute to V_2. Thus, V_2 is given completely by the scattered component of the dominant mode in (A9):

$$V_2=-\int\int_{\mathrm{ap}} dS'\mathbf{M}_2(\mathbf{r}')\cdot\mathbf{h}_2^{(1)}(\mathbf{r}'). \tag{A17}$$

In view of Eqs. (A17), (A16), (A14), and (A10), the total magnetic field in the half-space may be represented as follows:

$$\begin{aligned} \mathbf{H}_2(\mathbf{r})=\mathbf{H}_u(\mathbf{r})+\mathbf{H}(\mathbf{r})&=-I_2\mathbf{h}_2^{(1)}(\mathbf{r})\\ &\quad-\int\int_{\mathrm{ap}} dS'[Y_h(\mathbf{r},\mathbf{r}')\\ &\quad-\eta\mathbf{h}_2^{(1)}(\mathbf{r})\mathbf{h}_2^{(1)}(\mathbf{r}')]\cdot\mathbf{M}_2(\mathbf{r}'). \end{aligned} \tag{A18}$$

Reference to Eqs. (A9a) and the spherical mode expansion for Y_h in (A10) shows that the term subtracted from Y_h in (A18) is the real part of the dominant mode contribution to the kernel.

Determination of the Drag on a Cylinder at Low Reynolds Numbers

R. K. FINN
University of Illinois, Urbana, Illinois
(Received January 13, 1953)

Drag coefficients for a cylinder were measured at Reynolds numbers from 0.06 to 6.0 by observing the deflections of thin wires suspended in an air stream. The results confirm a theoretical equation proposed by Lamb and are consistent with experimental data of other investigators.

I. INTRODUCTION

IT is of theoretical and practical interest to evaluate the drag forces on a circular cylinder placed in a stream of fluid with its axis normal to the flow. For the region of viscous flow, a theoretical equation derived by Lamb[1] is generally used to predict the drag on a cylinder. It states that

$$C_D=\frac{8\pi}{\frac{1}{2}-\gamma-\ln(N_{Re}/8)}, \tag{1}$$

where γ is Euler's constant, $0.577\cdots$, N_{Re} is the Reynolds number based on the diameter of the cylinder, and C_D is the drag coefficient defined as

$$C_D=2Fg_c/\rho V^2D_eL, \tag{2}$$

in which F is the drag force expressed in weight units, g_c a Newton's law conversion factor, D_e the diameter of the cylinder, L the length of the cylinder, ρ and V the density and velocity of the fluid, respectively. Because of simplifying assumptions made in its derivation, the Lamb equation can be applied only when the Reynolds number is small compared to unity, i.e., at N_{Re} below about 0.1.

In contrast to the corresponding Stokes equation for spheres, the Lamb equation has not been verified by experiment. White[2] measured the terminal velocity of wires falling through liquids, but the drag coefficients which he calculated did not agree with predictions of the Lamb equation because wall effects became controlling in his experiments. At relatively high Reynolds numbers, the drag coefficients have been measured directly by a number of workers.[3–5] The data of Weiselsberger[4] extend to a Reynolds number as low as 4.1, but between 4.1 and 0.1, the upper limit of Lamb's equation, values of C_D are uncertain. Consequently, when presenting graphs of C_D *versus* N_{Re} for cylinders[5] the usual practice has been to draw a smooth curve connecting the data of Wieselsberger with the line representing the Lamb equation. Tomotika and Aoi[6] have recently refined the derivation of Lamb so as to allow the prediction of drag coefficients up to $N_{Re}=4$, but no data are available to confirm their calculations.

In the present investigation drag coefficients for a cylinder were measured at Reynolds numbers from 0.06 to 6.0. These measurements provide a direct test of Lamb's equation and also supply information about a flow region where neither theory nor experiment has been adequate.

II. EXPERIMENTAL

The experimental technique was similar in principle to that used earlier by Wieselsberger.[4] A fine wire with a weight attached to its lower end was suspended as a pendulum in a uniform stream of air flowing at a known rate. The drag force was calculated from the measured deflection of the pendulum. Since the weight itself hung outside the air stream, the deflection was due entirely to the drag on the cylindrical wire.

The flow chamber consisted of a short Lucite tube having an inside diameter of 1.90 cm. At the upstream end the flow chamber was connected to a standard flow nozzle and at the downstream end it was open to the atmosphere. In order to assure a uniform velocity profile past the test wire, it was suspended at a point only 2 mm downstream from a 200-mesh screen placed across the throat of the nozzle. Ranz[7] has shown that such an arrangement gives uniform velocity for a variety of flow conditions.

The top of the test wire was soldered onto a steel needle which served as a support and which passed through a small hole in the roof of the Lucite tube. The wire itself extended down through a hole in the floor of the tube and its lower end, weighted with a drop of solder, was protected from stray air currents by a small Lucite box. The essential parts of the apparatus are are shown in Fig. 1.

Deflections of the wire were observed with the aid of a microscope equipped with a 32 mm objective and a calibrated micrometer eyepiece. The point of observation was either at the floor of the Lucite tube or just outside the tube where the wire entered the Lucite box. From the deflection of the wire δ, the length exposed to the air stream L, and the weight of the solder ball W, it was possible to calculate the drag force from the

[1] H. Lamb, *Hydrodynamics* (Dover Publications, New York, 1945), sixth edition, p. 617.
[2] C. M. White, Proc. Roy. Soc. (London) **A186**, 472–479 (1946).
[3] E. F. Relf, Brit. ARC Repts. and Memoranda, No. 102 (1914).
[4] C. Wieselsberger, Physik. Z. **22**, 321 (1921).
[5] L. Schiller, *Handbuch der Experimentalphysik* (Akademische Verlagsgesellschaft, Leipzig, 1932), Vol. 4, Part 2, p. 337.
[6] S. Tomotika and T. Aoi, Quart. J. Mech. and Appl. Math. **3**, 140–161 (1950).
[7] W. E. Ranz, Ph. D. thesis in chemical engineering, University of Wisconsin (1950).

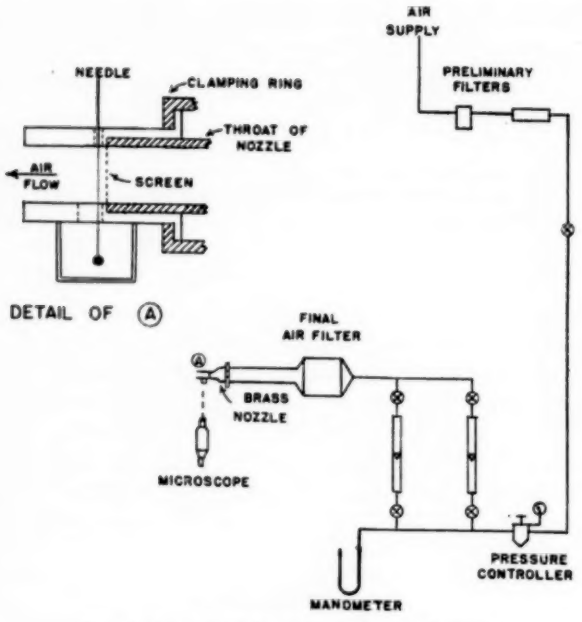


FIG. 1. Apparatus for measuring the drag of a fine wire in an air stream.

formula,

$$F = 2W\delta/L. \tag{3}$$

Equation (3) is based on the fact that the wires were flexible and therefore exerted no residual moment of force at the point of suspension.

Three sizes of wire were used consisting of tungsten, 4.1μ and 6.5μ in diameter, and of chromel, 12.6μ in diameter. The wires were examined under an electron microscope in order to determine their size and uniformity. The operations of cleaning, copper plating, and soldering the test wires were facilitated by the use of a jig which had been built to make probes for a hot wire anemometer.[8]

The range of air flows was from 20 to 730 centimeters per second, and the observed deflections varied from 36μ to a maximum of 1 mm. The weight of solder attached to the wire was in every case so much greater than the weight of the wire that the latter could be neglected.

III. RESULTS

The experimental data are shown in Fig. 2 along with the earlier data of Wieselsberger; also shown are the theoretical lines proposed by Lamb[1] and by Tomotika and Aoi.[6] The Lamb equation is confirmed for the region of strictly viscous flow. At higher Reynolds numbers the present data lie slightly below the predictions of Tomotika and Aoi, but there is a smooth fit with the experimental results of Wieselsberger.

The scatter of the data is not considered excessive

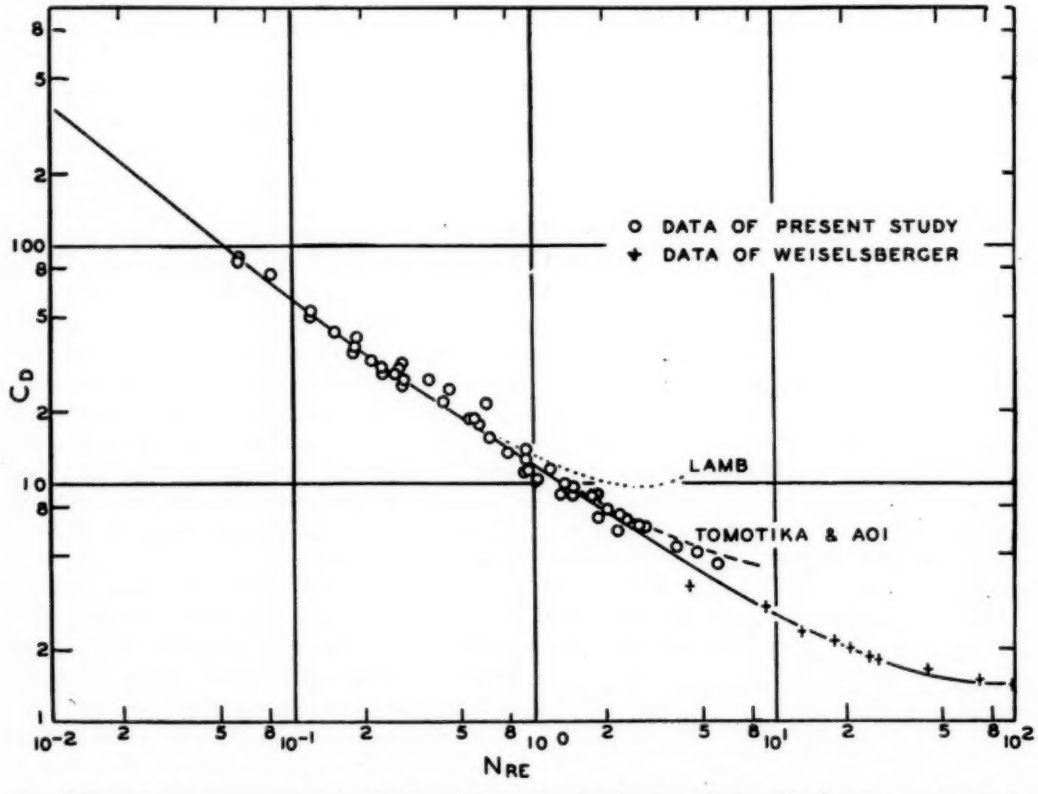


FIG. 2. Variation in drag coefficient with Reynolds number for circular cylinder. The line drawn through the points below $N_{Re}=0.5$ represents the theoretical Lamb equation, Eq. (1).

[8] R. W. Kunstman, Ph.D. thesis in chemical engineering. University of Illinois (1952).

in the light of experimental difficulties. Some of the abnormally high values for C_D may be accounted for by a slight kinking of the wire or by surface roughness, although efforts were made to avoid such interferences. The possible importance of kinking was illustrated in one series of tests in which the weight of the solder ball was insufficient to pull the wire taut. With the wire hanging as a loose coil, values of C_D were about 50 percent above the corresponding taut-wire values.

Since the diameter of the flow chamber was 1500 times the diameter of the largest wire used, any influence of the wall can be considered negligible. White[2] has shown that the magnitude of the wall effect depends upon the Reynolds number, and presumably an interference would have been noted if the present measurements had been extended to values of N_{Re} below 0.01. Inherent limitations in the technique prevented investigation of the drag at such low Reynolds numbers, however.

IV. ACKNOWLEDGMENTS

The author wishes to thank W. E. Ranz for his many helpful suggestions. This work was performed in the Engineering Experiment Station at the University of Illinois as part of a contract for aerosol studies, Contract No. At (30-3)-28, United States Atomic Energy Commission, H. F. Johnstone, technical director. Further details and a tabulation of the data may be found in Technical Report No. 8 on the contract, obtainable from the Technical Information Service, United States Atomic Energy Commission, Oak Ridge, Tennessee.

NOMENCLATURE

C_D = drag coefficient, dimensionless = $(2Fg_c)/(\rho V^2 D_c L)$.
D_c = diameter of cylinder, cm.
F = drag force, g force.
g_c = conversion factor = 980.6 (g mass/g force) (cm/sec^2).
L = length of cylinder, cm.
$N_{Re} = (D_c V \rho)/\mu$ = Reynolds number, dimensionless.
V = average velocity of fluid, cm/sec.
W = weight attached to cylinder, g force.
γ = Euler's constant = 0.5772 $\cdots$, dimensionless.
δ = deflection, cm.
μ = fluid viscosity, poise = g mass/cm, sec.
ρ = fluid density, g mass/cm^3.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

Charged Disk in Cylindrical Box

W. R. SMYTHE
Norman Bridge Laboratory, California Institute of Technology, Pasadena, California
(Received February 25, 1953)

The mixed boundary value problem of the potential inside a cylindrical conducting box of radius a and height $2c$ in which is mounted coaxially and concentrically a thin flat charged conducting disk of radius b is solved approximately. A formula for the coefficients in the Bessel function potential series is derived which is valid for $c > \frac{1}{4}a$ and $b < \frac{3}{4}a$. The boundary conditions are met rigorously except that the disk is not quite flat. Figures show its deviation for twelve parameter values. A formula for narrow gaps is derived. Upper and lower limits for the capacitance for twelve parameter values are tabulated.

IN a previous paper[1] the mixed boundary value problem of a nearly flat disk electrode applied coaxially to the end of a right circular cylinder of resisting material was solved. A very similar problem whose solution is often needed is that of finding the potential inside a circular conducting cylindrical box of radius a and height $2c$ in which is mounted coaxially and concentrically a thin flat circular charged conducting disk of radius b. The upper half of this box is shown in Fig. 1. As in the first case the solution found satisfies all conditions rigorously except that the disk is not quite flat.

A solution of Laplace's equation in the cylindrical coordinate system ρ, ϕ, z which gives $V=0$ on the walls of the box is

$$V = \sum_{n=1}^{\infty} A_n \sinh(k_n c - |z|) J_0(k_n \rho), \tag{1}$$

provided that k_n is chosen from the tables[2] so that

$$J_0(k_n a) = 0. \tag{2}$$

As in the first problem,[1] a uniform charge distribution over the disk area is combined with that on a freely charged disk in such a ratio that the resultant potential over the disk area is nearly constant. Thus when $0 < \rho < b$ we have

$$\sigma = -\epsilon \frac{\partial V}{\partial z} = \frac{Q[(b^2-\rho^2)^{-\frac{1}{2}} + D]}{2\pi b(2+bD)}, \tag{3}$$

[1] W. R. Smythe, J. Appl. Phys. 24, 70 (1953); MKS units are used throughout.

[2] H. T. Davis and W. J. Kirkham, Bull. Am. Math. Soc. 33, 760 (1927). These tables give both $k_n a$ and $J_1(k_n a)$.

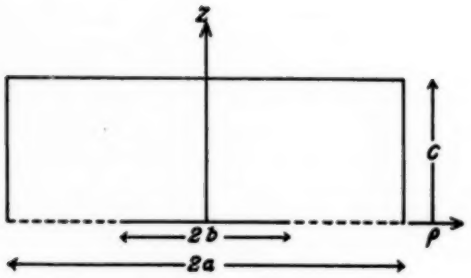


FIG. 1. Upper half of cylindrical box. Electric line of force is dotted.

where σ is the density on one face of the disk, Q is the total charge including both faces, and D is an adjustable constant. We now differentiate Eq. (1) with respect to z, set $z=0$, equate it to Eq. (3) when $\rho<b$ and to zero when $b<\rho<a$, and determine A_n in the standard way.[3] The integrals are the same as before,[1] so the result is

$$A_n=\frac{Q[\sin(k_nb)+bDJ_1(k_nb)]}{\pi b\epsilon(2+bD)k_n^2a^2\cosh(k_nc)J_1^2(k_na)}. \quad (4)$$

As before[1] D is evaluated by putting $z=0$ and $c=\infty$ in Eqs. (1) and (4) and summing the $\sin(k_nb)$ terms and the $J_1(k_nb)$ terms of the series separately for the ρ-values 0, $\frac{1}{4}b$, $\frac{1}{2}b$, $\frac{3}{4}b$, and b. The difference in these five values of V_0 is then minimized by adjusting D. This is not difficult for sometimes two sets of values for which D differs by 20 percent are almost equally good. Two curves are then drawn, one showing maximum permissible values of D as a function of b/a and the other minimum values. The analytical curve $bD=-0.58(b/a)^{3.5}$ lies well between them. When $c\neq\infty$, it is clear that bD must be modified to compensate for the variation of potential over the disk area produced by its images in the top and bottom walls of the box. The difference in the image potentials at $\rho=0$ and $\rho=b$ is proportional, from Eq. (1), to $1-J_0(k_nb)$ or to $k_n^2b^2$. The strength of the image itself is proportional, from Eq. (4), to $\sin(k_nb)$ or $J_1(k_nb)$. Thus the b dependence of bD will probably involve a third-order combination of k_1b and $\sin(k_1b)$ or $J_1(k_1b)$. Since, roughly, k_1a is 2.4 and k_2a is 5.5, the first two terms of the potential of the first image at $z=\pm2c$ are proportional to $\exp(-4.8c/a)$ and $\exp(-11c/a)$, and the first term of the second image to $\exp(-9.6c/a)$. The last two terms average roughly $\exp(-10c/a)$. By testing forms suggested by such reasoning the following formula was arrived at

$$bD=-0.58(b/a)^{3.5}+(b/a)J_1^2(2.4b/a)\times[19.5e^{-4.8c/a}+60e^{-10c/a}]. \quad (5)$$

The 19.5 and 60 factors were chosen to minimize the potential variation over the disk at five values, 0, $\frac{1}{4}a$, $\frac{1}{2}a$, $\frac{3}{4}a$, and a of b and three values a, $\frac{1}{2}a$, and $\frac{1}{4}a$ of c. The same sums can be used in evaluating these potentials as for $c=\infty$ except that the first few terms are modified by the $\tanh(k_nc)$ factor.

Seven places were carried in the above calculations. This is the limit imposed by linear interpolation in the tables.[4] Terms up to about $n=12$ could be checked only by repeating the calculation independently. The smaller large n terms could be checked rapidly by taking the ratio on a slide rule of successive terms having the same phase. Errors in the sixth or seventh place were easily located by abrupt changes in this smoothly varying ratio. This checks all previous operations including the use of the tables. The sums were found as already indicated[1] and their accuracy can be estimated by comparing sums with different term groupings. They appeared accurate to six places.

If $V(\rho)$ is the potential given by Eq. (1) at $z=0$ and if V_0 is the potential at some chosen point on $z=0$, then the displacement of the V_0 equipotential from the $z=0$ plane at ρ is, from Eqs. (1) and (3),

$$\Delta z=\frac{V-V_0}{\partial V/\partial z}=\frac{b(V-V_0)}{[(1-\rho^2b^{-2})^{-\frac{1}{2}}+bD]V'}, \quad (6)$$

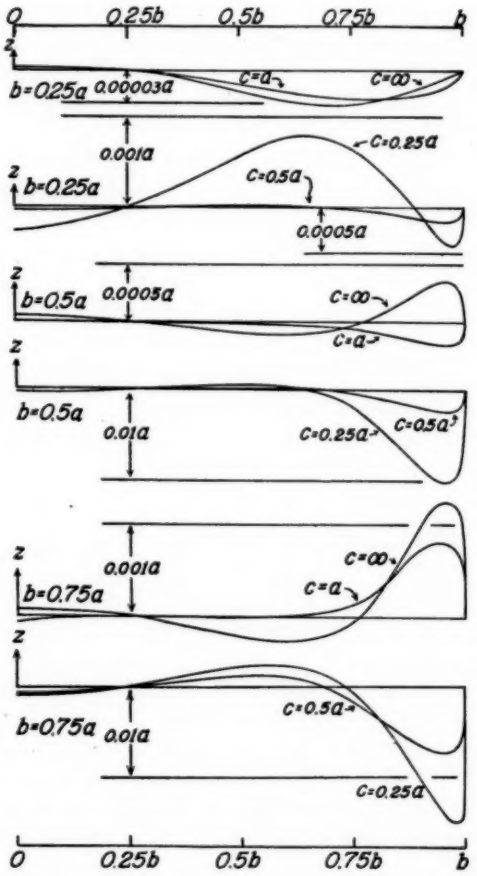


FIG. 2. Radial sections of the equipotential surface that passes through $\rho=\frac{1}{4}b$. The vertical scale is greatly expanded as indicated to show the deviation from a flat disk.

[3] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), p. 177.

[4] The Staff of the Computation Laboratory, *Tables of Bessel Functions* (Harvard University Press, Cambridge, 1947).

where V' is a unit potential. This deviation at the twelve check points is shown in Fig. 2. It will be noted that the disk is flattest when b is small and c is large. The maximum deviation of the equipotential passing through $\frac{1}{4}b$ and the edge given by the above formulas is shown as a function of b and c in Fig. 3. These deviations could probably be reduced to 0.1 or 0.01 of their present value by inserting a ρ^2 correction term in Eq. (3) as was done in the previous work.[1]

NARROW GAP FORMULA

It is clear from Fig. 3 that when the disk almost touches the wall, so that $b\rightarrow a$, the A_n given by Eq. (4) is very inaccurate. This case can be handled by making the potential rigorously V_0 on the disk and estimating its value in the gap. The latter must be V_0 at $\rho=b$, 0 at $\rho=a$ and must give E proportional to $(b^2-\rho^2)^{-\frac{1}{2}}$, when $b<\rho<a$ and $\rho\rightarrow b$. Furthermore, it should give the known rigorous result when $b=a$. Such a potential is

$$V=V_0\left[1-\frac{\cos^{-1}(b/\rho)}{\cos^{-1}(b/a)}\right] b<\rho<a. \tag{7}$$

With $z=0$, Eq (1) is equated to V_0 when $\rho<b$ and to (7) when $b<\rho<a$, and the coefficient A_n is found[3] after repeated integration by parts to be

$$\frac{2b(a^2-b^2)^{\frac{1}{2}}\operatorname{csch}(k_nc)V_0}{k_na^3J_1^2(k_na)\cos^{-1}(b/a)} \times\sum_{s=0}^{\infty}\frac{[k_na(1-b^2a^{-2})]^s}{(2s+1)!!}J_{s+1}(k_na). \tag{8}$$

When $b=a$, only the first term survives, and this becomes

$$A_n=\frac{2\operatorname{csch}(k_nc)}{k_naJ_1(k_na)}V_0, \tag{9}$$

which is the rigorous value for an infinitely narrow gap. Large values of n are needed when z is small, and for such values Eq. (8) is a poor formula for numerical calculation unless the gap is very narrow. It is interesting because, if δ is the distance from the disk edge, it

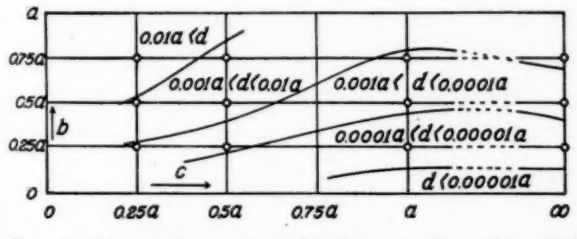


FIG. 3. The maximum deviation from a plane disk of the equipotential surface that passes through $\rho=\frac{1}{4}b$ as given by Eqs. (1) and (4) over a wide range of values of b and c. Check points are shown by circles.

TABLE I. Limits of capacitance (MKS units).

b	c	Minimum	Maximum	Difference	Mean
0.25a	∞	2.3253 ϵa	2.3256 ϵa*	0.0003 ϵa	2.3255 ϵa
0.50a	∞	5.6145 ϵa*	5.6358 ϵa	0.0213 ϵa	5.6252 ϵa
0.75a	∞	11.140 ϵa*	11.295 ϵa	0.1548 ϵa	11.217 ϵa
0.25a	a	2.3444 ϵa	2.3449 ϵa*	0.0004 ϵa	2.3446 ϵa
0.50a	a	5.6964 ϵa*	5.7100 ϵa	0.0136 ϵa	5.7032 ϵa
0.75a	a	11.325 ϵa*	11.409 ϵa	0.0836 ϵa	11.367 ϵa
0.25a	0.5a	2.5419 ϵa	2.5494 ϵa*	0.0075 ϵa	2.5456 ϵa
0.50a	0.5a	6.4790 ϵa	6.6188 ϵa*	0.1398 ϵa	6.5489 ϵa
0.75a	0.5a	12.663 ϵa	13.493 ϵa*	0.8306 ϵa	13.078 ϵa
0.25a	0.25a	3.1986 ϵa	3.2433 ϵa*	0.0447 ϵa	3.2209 ϵa
0.50a	0.25a	8.9586 ϵa	10.152 ϵa*	1.1933 ϵa	9.5552 ϵa
0.75a	0.25a	17.720 ϵa	22.123 ϵa*	4.403 ϵa	19.922 ϵa

covers the transition from the δ^{-1} infinity in E that occurs when $b=a$ to the $\delta^{-\frac{1}{2}}$ infinity that appears in E for gap widths not zero. The range of c covered by Eq. (8) should be from about a–b to ∞.

CAPACITANCE

Upper and lower limits for the capacitance between disk and box can now be found. Consider the top half of the box as shown in Fig. 1. Almost invariably the potential given by Eqs. (1) and (4) when $\rho\lesssim b$ will be a maximum or minimum at $\rho=b$. If it is a maximum then the equipotential V_M passing through the edge lies below the $z=0$ plane when $\rho<b$, so the capacitance between this equipotential and the walls is less than that of a flat disk. On the other hand, the minimum potential V_m given by Eqs. (1) and (4) at some $\rho_m<b$ lies elsewhere above the $z=0$ plane, except where it hooks over the edge and meets this plane normally at $\rho>b$. Thus it entirely encloses the flat disk, and the capacitance between it and the walls exceeds that of the disk. A similar argument can be applied when V_m passes through the edge. Thus

$$(Q/V_M)<C<(Q/V_m), \tag{10}$$

where V_m and V_M are the minimum and maximum values of V at $z=0$, $\rho\lesssim b$ given by Eqs. (1) and (2). The limits of capacitance obtained in this way at the twelve check points appear in Table I.

In most cases the potential is nearly constant out to $\rho=\frac{1}{2}b$ or $\rho=\frac{3}{4}b$ and then increases or decreases rapidly to its value at the edge. The edge value limit is starred in the table. Because of the greater area and charge density for values of ρ near the edge, the potential there carries more weight which should counterbalance somewhat the smaller range of ρ to which it apples. It is believed, therefore, that the mean value given in the last column of the table is close to the true value and probably never differs from it by as much as one-fourth of the difference given in the third column. For narrow gaps Eq. (8) gives A_n, and limits for the capacitance are given by Q/V, where for one limit Q is the charge on the disk and for the other the charge on the box.

The Darlington Problem*

LOUIS WEINBERG
Hughes Research and Development Laboratories, Culver City, California
(Received February 2, 1953)

Any RLC transfer function with an even or odd numerator is realized as the transfer voltage ratio of a Darlington network, where such a network is defined as one containing lossless elements plus only *one* resistance. The single resistance appears *within* the network, and the termination is thus lossless. An unbalanced network that contains no mutual inductance is achieved by the procedure. The method makes use of Norton's theorem, zero shifting, and the paralleling of ladder networks. The number of ladders and elements required in the realization is kept small by a technique demonstrated in the paper.

INTRODUCTION

DARLINGTON made the important contribution[1,2] of demonstrating that any positive real function is realizable as the input impedance of a lossless two terminal-pair network terminated in resistance. This proof forms the basis of a synthesis procedure for transfer functions of such networks. The Darlington procedure, however, requires ideal transformers in general, and the synthesis of the transfer characteristic is for *magnitude alone, not for magnitude and phase.* Another way of specifying this restriction is by stating that the transfer function can be realized for both magnitude and phase provided that the numerator is either an even or odd function of the complex variable.

In this paper a procedure is presented for realizing such transfer functions, i.e., functions whose numerator is either even or odd with positive coefficients, as the transfer voltage ratio of an unbalanced Darlington network. As generally used, the term Darlington network means a lossless coupling network terminated in resistance; in the present paper the term is used in a more general sense to mean a network of lossless elements plus *one* resistance: the resistance is necessary to make the transfer function have poles off the j axis, but it may exist *within the network* or as a termination. The network achieved here requires no mutual inductance and has a reactance termination. In a subsequent paper the same problem will be treated by another procedure which works through the intermediate step of realizing a lattice in order to achieve the final unbalanced network.

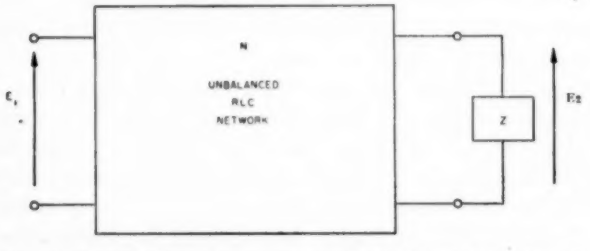


FIG. 1. Desired form of network in which N contains only one resistance and Z is a pure reactance.

The problem of this paper, namely, the realization of a transfer voltage ratio by means of a Darlington network, can already be solved by means of a procedure analogous to the Guillemin one for RC synthesis;[3] but in this paper we accomplish for RLC synthesis what was previously done for RC synthesis,[4] that is, we demonstrate a method that makes use of Norton's theorem and reduces the number of elements and ladders required in the final network. It will be assumed that the reader is familiar with the last-cited reference; in that paper are given the general formulas and techniques upon which the present discussion will be based. For purposes of clarity and definition of symbols, we shall restate the general formulas in this paper.

STATEMENT OF THE PROBLEM

The given transfer function

$$K=p(s)/q(s), \tag{1}$$

where p is an even or odd polynomial with positive coefficients and q is a Hurwitz polynomial, is to be realized within a multiplicative constant H as the transfer voltage ratio of an unbalanced Darlington network containing no mutual inductance. The form of the desired network is shown in Fig. 1.

The network to be synthesized may be divided into two parts, as shown in Fig. 2. It has been shown[4] that this network is characterized by

$$K\equiv\frac{E_2}{E_1}=\frac{y_{12a}K_b}{y_{22a}+Y_b}, \tag{2}$$

where

$$\left.\begin{aligned} y_{12a}&=I_{sc}/E_1\\ K_b&=E_2/E\end{aligned}\right\}, \tag{3}$$

I_{sc} is the current obtained in a short-circuit across

* This paper is based on a chapter of Technical Report No. 201, Research Laboartory of Electronics, Massachusetts Institute of Technology, September, 1951. The research was conducted under the supervision of Dr. E. A. Guillemin and was supported in part by the Air Materiel Command, The Army Signal Corps, and the U. S. Office of Naval Research.

[1] E. A. Guillemin, *A Summary of Modern Methods of Network Synthesis*, Advances in Electronics, Vol. III (Academic Press, Inc., New York, 1951).

[2] S. Darlington, J. Math. and Phys. **18**, 257–353 (1939).

[3] E. A. Guillemin, J. Math. and Phys. **28**, 22–42 (1949).

[4] L. Weinberg, J. Appl. Phys. **24**, 207 (1953).

terminals a–b, Y_b is the admittance seen looking to the right at a–b, and y_{22a} is the short-circuit driving-point impedance of network N_a.

Each of the partitioned networks may be realized separately. To realize N_a, we may parallel ladder structures, each ladder being obtained by LC zero-shifting techniques[1] which start from the right and work toward the left. For N_b (i.e., for N_b' and Z), on the contrary, we work from left to right and must use *only one ladder*. Just as only one RC ladder is necessary to realize a transfer voltage ratio with all negative real zeros, so only one ladder is needed for the LC network N_b when all its zeros are on the j axis. However, when the transfer function contains one or more quadruplets of zeros, i.e., a pair of complex conjugate zeros in the left half-plane plus a pair in the right half-plane symmetrically placed with respect to the origin, then more than one ladder is needed. It is therefore necessary that only zeros on the j axis be associated with the transfer function of N_b. The distribution of poles to the transfer functions of N_a and N_b determines the distribution of the partial fractions of the positive real denominator. It is necessary (and always possible) to make each ladder structure of N_a end

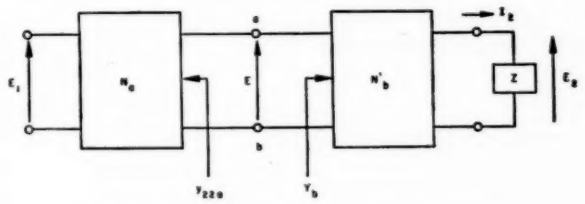


FIG. 2. Network after division into two parts.

on the left with a series branch, while the ladder for N_b must end on the right with a shunt branch.

It is thus clear that the synthesis will be accomplished if we first show how to modify the polynomials in Eq. (1) so that they can be identified with the driving-point and transfer characteristics of the two partitioned networks given by Eq. (2). Then, after the identifications have been made, the networks N_a and N_b may be developed. Since the former is a group of paralleled ladders, a method is shown for making the required number of ladders small. Each of the component ladders may be associated with only a part of y_{22a} in order to use fewer elements. This technique can be used when, in decomposing the transfer function of N_a, we bring about zero-pole cancellations; the details of the technique are the same as for RC synthesis.[4]

THE SYNTHESIS PROCEDURE PRESENTED

Let the q of Eq. (1) be defined by

$$q=m_2+n_2, \tag{4}$$

where m_2 is the even and n_2 the odd part of q. If p is even we divide the numerator and denominator of K by n_2; if it is odd, we divide by m_2. Suppose without loss of

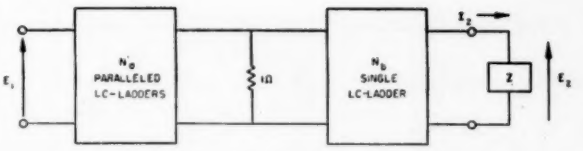


FIG. 3. Final form of Darlington network where the termination Z is composed of lossless elements.

generality that p is an even polynomial m_1. Therefore,

$$K=\frac{m_1}{m_2+n_2}=\frac{m_1/n_2}{1+(m_2/n_2)}. \tag{5}$$

We now make identifications with the terms of Eq. (2) so that a small number of ladders will be required in N_a, each ladder being characterized by a transfer function with a low degree denominator. The required identifications are given by

$$\left.\begin{aligned} y_{12a}&=m_{1a}/n_{2a}\\ y_{22a}&=1+(m_{2a}/n_{2a}) \end{aligned}\right\} \tag{6}$$

and

$$\left.\begin{aligned} K_b&=m_{1b}/m_{2b}\\ Y_b&=n_{2b}/m_{2b} \end{aligned}\right\}, \tag{7}$$

where by definition

$$\left.\begin{aligned} m_1&\equiv m_{1a}m_{1b}\\ n_2&\equiv n_{2a}m_{2b}\\ \frac{m_2}{n_2}&\equiv\frac{m_{2a}}{n_{2a}}+\frac{n_{2b}}{m_{2b}} \end{aligned}\right\}. \tag{8}$$

It is important to note that m_{1b} must be chosen as a constant or as an even polynomial with zeros only on the j axis (since N_b must be a single ladder). In addition, the odd part of n_2 (i.e., its zero at the origin) must be associated not with K_b but with y_{12a}, because y_{12a}, a transfer admittance, must be an *even polynomial over an odd* (or an *odd over even*), while K_b, a dimensionless voltage ratio, must be an *even over an even polynomial*. Finally, since q is a Hurwitz polynomial, m_2/n_2 is a positive real reactance function[5] whose partial fractions are also positive real and hence separately realizable. Thus the functions m_{2a}/n_{2a} and n_{2b}/m_{2b} are positive real sums of the partial fraction components of m_2/n_2.

In the realization of N_a the constant term is immediately removed from y_{22a} as a shunt branch, so that the remaining synthesis embodying paralleling of ladders and zero shifting need be done with a two-element kind network,[1] and is analogous to the procedure used for RC synthesis.[4] The final network has the form shown in Fig. 3.

It was shown in the RC synthesis paper[4] that it is possible to reduce the number of ladders from that

[5] E. A. Guillemin, *The Mathematics of Circuit Analysis* (John Wiley and Sons, Inc., New York, 1949).

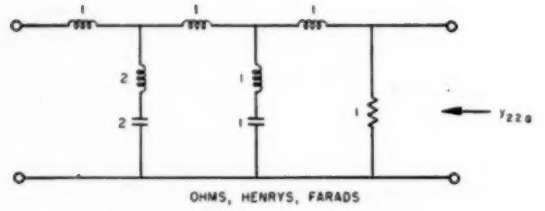


FIG. 4. Network realization for N_a.

obtained by using two successive terms of p as tht numerator of one ladder's transfer function. A similar procedure applies here with the differences pointed out below.

If the numerator is originally odd, we factor out an s to obtain an even polynomial $f(s^2)$; if the numerator is even, we consider the total numerator as $f(s^2)$. The substitution $y=s^2$ gives $f(y)$, a polynomial with no missing terms and of degree n. We then apply the Sturm test[6] to determine the number of negative real roots of $f(y)$; designate this number by r. Finally, with these new definitions, the formula for the required number of ladders is the same as that given by Eq. III.5 of reference 4, namely, the smallest integer t satisfying the relationship

$$t \geqslant (n-r+1)/2. \tag{9}$$

As an illustration, consider

$$p_a=s^8+17s^6+121s^4+455s^2+750, \tag{10}$$

which yields

$$f(y)=y^4+17y^3+121y^2+455y+750. \tag{11}$$

The Sturm test shows this to have two negative real roots, and we then find $f(y)$ in factored form to be

$$f(y)=(y^2+6y+25)(y+5)(y+6), \tag{12}$$

which roots are needed for the subsequent synthesis. Since $n=4$ and $r=2$, the required number of ladders in N_a is 2. Using the straightforward procedure of taking two successive terms at a time would have required three ladders for the network realization.

The roots of the total denominator, it is finally pointed out, need not be found, thus saving the designer from one of the tedious calculations of many synthesis procedures.

ILLUSTRATIVE EXAMPLE

An example is chosen in which the zeros lie on the j axis. Thus no paralleling of ladders will be required for the realization of N_a. The technique of paralleling ladders, which is necessary for a quadruplet of zeros, is adequately illustrated in the analogous RC example[4] given in the reference.

The given transfer function is

$$K=\frac{(s^2+0.25)(s^2+1)}{26s^9+94s^8+151s^7+334s^6+264s^5+298s^4+141s^3+80s^2+18s+6}.$$

The even and odd parts of the denominator are given by inspection as

$$m_2=26s^9+151s^7+264s^5+141s^3+18s,$$
$$n_2=94s^8+334s^6+298s^4+80s^2+6.$$

After division of numerator and denominator by n_2, it is necessary to decompose m_2/n_2 into two parts, each of which is the sum of appropriately chosen partial fraction components. This breakdown is computed as

$$\frac{m_2}{n_2}=\frac{m_{2a}}{n_{2a}}+\frac{n_{2b}}{m_{2b}}=\frac{16s^4+9s^2+1}{26s^5+21s^3+3s}+\frac{3s^3+7s}{s^4+5s^2+6}.$$

We can now make identifications of the terms in Eq. (2) as follows:

$$y_{12a}=\frac{(s^2+0.25)(s^2+1)}{26s^5+21s^3+3s},$$

$$y_{22a}=1+\frac{16s^4+9s^2+1}{26s^5+21s^3+3s},$$

and

$$K_b=\frac{1}{s^4+5s^2+6},$$

$$Y_b=\frac{3s^3+7s}{s^4+5s^2+6}.$$

We first realize the ladder N_a. The conductance of one mho is removed as an initial shunt branch. It is now necessary to develop the remainder of y_{22a}, i.e., $y_{22a}'=y_{22a}-1$, simultaneously zero shifting to insert the zeros of y_{12a} at $s=\pm j0.5$ and $s=\pm j1$. This is accomplished as in Sec. XXII of the first reference. (We merely let

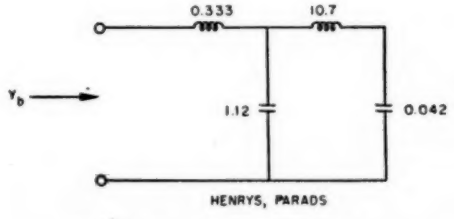


FIG. 5. Network realization for N_b.

z_{11} of that paper be y_{22a}' and z_{12} be y_{12a}. The network obtained is the dual of the one shown there in Fig. 15 with the terminal-pairs interchanged.) The network N_a is shown in Fig. 4.

Ladder N_b must have an input admittance given by N_b and a transfer function with all its zeros at infinity. This is achieved simply by a continued fraction expansion of Y_b that yields series inductance arms and shunt capacitance arms. We therefore first invert Y_b and then perform the divisions to obtain a continued fraction. Thus,

[6] L. E. Dickson, *New First Course in the Theory of Equations* (John Wiley and Sons, Inc., New York, 1939).

$$
\begin{array}{l}
\quad\quad\quad

from the purely viscous component of the deformation and viscoelastic effects are absent. They can be considered to be Newtonian since they were either determined at low rates of shear or extrapolated to zero shear stress. The experimental facts concerning viscosity-temperature characteristics are as follows. The activation energy, E, is high at low temperatures, varies quite rapidly over a certain temperature range, and approaches a considerably lower value at high temperatures. It is to be noted that this type of behavior is not explainable by a simple superposition of two rate processes as in that case the process of low E would predominate at low temperatures. Thus, if the rates of shear attributable to two different mechanisms are added, one obtains

$$\frac{1}{\eta}=\frac{1}{\eta_1}e^{-E_1/RT}+\frac{1}{\eta_2}e^{-E_2/RT},$$

where η is the viscosity. It is easily shown from this relation that the process of lower E dominates at low temperatures.

An equation of the form

$$\frac{\eta}{T}=Ae^{B/T}[1+ce^{D/T}],$$

was proposed by Douglas[5] to describe the viscosity-temperature characteristics of glasses. The term in brackets was interpreted as arising from the change of atomic configuration with temperature. At about the same time the writer proposed independently[6] the relation

$$\eta=\eta_0 e^{E_0/RT}[1+ke^{E_a/RT}], \tag{1}$$

to describe the viscosity-temperature characteristics of high polymers. Again the quantity in brackets was interpreted as due to changes in the configurations of the molecules with temperature. Careful examination of recent data[3] for glasses indicated, however, that these equations are not quite satisfactory in the transition region since they generally predict a faster change of E with temperature than is experimentally observed.

An empirical equation of the form

$$\log\eta=A+[B/(T+C)], \tag{2}$$

has been used quite successfully for a long time in the glass industry.[7] Gutmann and Simmons[8] have recently derived this equation in a rather formal way. If the variation of E with temperature is written as

$$E=E_0+f(T), \tag{3}$$

and $f(T)$ is expressed as a convergent power series in T, then Eq. (2) results from the relation $\eta=$const. exp (E/RT), when a certain number of terms are retained in the series. This procedure is formally quite satisfactory but the adjustable parameters are difficult to interpret physically.†

In this note, an attempt is made to determine the form of $f(T)$ on the basis of a physical argument using the concept of short-range order and its variation with temperature. The final relation turns out to be a generalization of the above equations.

TABLE I. Constants of Eq. (10) fitted to viscosity-temperature data on various materials.

Material	$\log_{10}\eta_0$, poises	T_0 in °K	E K cal/mole	U K cal/mole	Avg. deviation in $\log_{10}\eta$
Glasses[a]					
No. 2	−2.135	560	18.28	11.77	0.073
No. 10	−1.446	560	4.57	17.11	0.029
No. 330	−2.249	562	18.28	10.47	0.002
Polyisobutylene[b]					
80 000 MW	−0.611	170	6.03	1.86	0.038
660 000 MW	+2.730	190	5.03	1.66	0.036
Phenol-formaldehyde novolak resin[c]					
Resin *B*					
MW=471	−2.705	305	0.914	1.737	0.006
Resin *C*					
MW=587	−3.014	320	2.742	1.508	0.011
Water[d]	−0.5907[e]	200	1.782	0.237	0.022[f]

[a] Data from reference 3.
[b] Data from reference 2.
[c] Data from reference 4.
[d] Data from reference 13.
[e] Viscosities in millipoises.
[f] Deviation in η.

[5] R. W. Douglas, J. Soc. Glass Technol. **33**, 138 (1949).
[6] G. J. Dienes, Paper presented at 1950 Meeting of the Society of Rheology. See Rheol. Bull. **19**, No. 2, **2** (1950).

II. THEORETICAL CONSIDERATIONS

As the temperature is varied different types of structural changes may take place in a liquid (for example association in water and glasses, configurational changes in high polymers, etc.). In general, one certainly expects the degree of order, i.e., the regularity of atomic positions, to decrease as the temperature is raised. Further, it is generally accepted that viscous flow involves configurational changes in the liquid.[9] Consequently, the less ordered a liquid is the easier it is to bring about the configurational changes required for flow. Thus, one expects viscosity to decrease as the degree of order is reduced. Mack[10] has developed this dea in some detail. His results suggest that one may assume the fluidity to be proportional to the degree of disorder. On this basis one may write

$$\phi=\phi'(1-\sigma), \tag{4}$$

where $\phi=$fluidity and $\sigma=$degree of short-range order.

[7] G. S. Fulcher, J. Am. Ceram. Soc. **8**, 339 (1925).
[8] F. Gutmann and L. J. N. Simmons, J. Appl. Phys. **23**, 977 (1952).
† It may be remarked that the introduction of a distribution of activation energies would be formally quite similar to Gutmann and Simmons' treatment with similar difficulties of physical interpretation. Whenever time-dependent phenomena are involved however, the use of a distribution of activation energies appears appropriate. See, for example, the recent paper by W. J. Lyons, J. Appl. Phys. **24**, 217 (1953).
[9] For a recent review see B. S. Harrap and E. Heymann, Chem. Revs. **48**, 45 (1951).
[10] C. Mack, J. Appl. Phys. **19**, 1082 (1948).

In terms of viscosity Eq. (4) becomes

$$\eta=\eta'\left[\frac{1}{1-\sigma}\right], \tag{5}$$

where η = viscosity, η' = viscosity of completely disordered material, i.e., at $\sigma=0$. η' is expected to have the normal exponential dependence on temperature, i.e.,

$$\eta'=\eta_0 e^{E/RT} \tag{6}$$

where E is the activation energy for viscous flow in the disordered liquid ($\sigma=0$). However, since σ is a function of temperature another temperature dependent term arises from the quantity in the brackets in Eq. (5). Frenkel[11] gives a simple relation for short-range order as a function of temperature which should be quite satisfactory at elevated temperatures. The relation is

$$\sigma=\tanh(U/2RT),$$

where U is the energy of ordering. In the present application, following Eyring,[12] U may be called the "structure activation energy." As written, this equation would give $\sigma=1$ at $T=0$. In order to obtain $\sigma=1$ at a finite temperature T_0 an obvious modification of the above equation is

$$\sigma=\tanh[U/2R(T-T_0)]. \tag{7}$$

Combination of (6) and (7) gives, on the basis of this picture, the complete temperature dependence of viscosity as

$$\eta=\eta_0 e^{E/RT}\left[1\Big/1-\tanh\frac{U}{2R(T-T_0)}\right]. \tag{8}$$

By expressing tanh as the exponential and carrying out some algebraic reductions Eq. (8) may be reduced to

$$\eta=\frac{\eta_0}{2}e^{E/RT}\left[e^{U/R(T-T_0)}+1\right]. \tag{9}$$

When U is zero or T is very large Eq. (9) reduces to the simple exponential form, Eq. (6). When T_0 is small relative to T the two-exponential relation, Eq. (1), is obtained. When U is large compared to E and the exponential in the brackets is large compared to unity the simple logarithmic relation of Eq. (2) is obtained. Equation (9), therefore, appears to be a generalization of the earlier viscosity-temperature relations and the constants of this equation have well-defined physical meaning.

III. COMPARISON WITH EXPERIMENT

It is much easier to fit Eq. (9) to experimental data if unity in the brackets may be neglected relative to the exponential. This has been found to be justified in the cases to be discussed but may not be generally permissible, particularly at high temperatures. With this approximation, Eq. (9) becomes, in logarithmic form,

$$\log_{10}\eta=\log_{10}\frac{\eta_0}{2}+\frac{E}{2.3RT}+\frac{U}{2.3R}\frac{1}{T-T_0}. \tag{10}$$

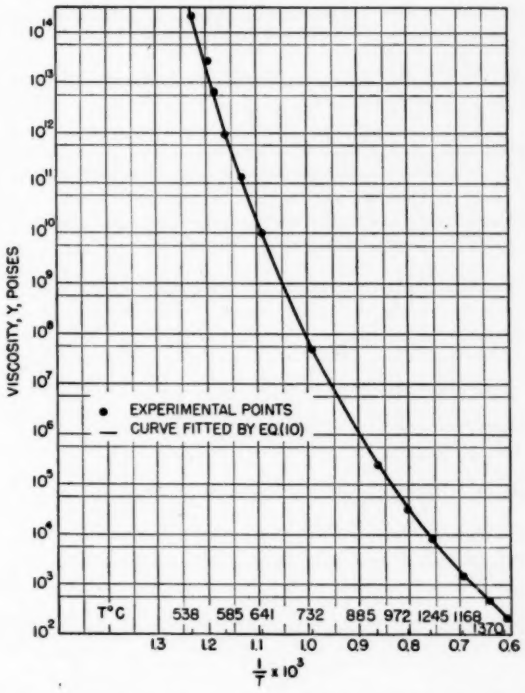

VISCOSITY, η, POISES
EXPERIMENTAL POINTS
CURVE FITTED BY EQ(10)
T°C
538 585 641 732 885 972 1245 1168 1370
1.3 1.2 1.1 1.0 0.9 0.8 0.7 0.6
1/T x 10³

FIG. 1. Comparison of experimental data and Eq. (10) for glass No. 10 studied by Robinson and Peterson (reference 3).

Curve fitting was accomplished by means of the following steps:

(a) Differentiate Eq. (10) with respect to $1/T$. The slope m is then given by

$$m=\frac{d(\log\eta)}{d(1/T)}=\frac{E}{2.3R}+\frac{U}{2.3R}\frac{T^2}{(T-T_0)^2}. \tag{11}$$

Then m *vs* $(T/T-T_0)^2$ was plotted. This plot will be linear for proper choice of T_0. From the intercept and the slope E and U were determined. m was obtained from the experimental data by a point by point calculation of the slope.

(b) With the above values of E, U, and T_0, η_0 was calculated from the experimental $\eta-T$ data and averaged.

(c) Assuming that E and η_0 are correct, more precise values for U and T_0 were obtained by plotting

$$y=\log\eta-\log\frac{\eta_0}{2}-E/RT \quad vs \quad 1/(T-T_0).$$

This plot is linear for the correct choice of T_0; the line goes through zero, and its slope is $U/2.3R$.

[11] J. Frenkel, *Kinetic Theory of Liquids* (Oxford University Press, London, 1946), pp. 62–66.

[12] Glasstone, Laidler, and Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., 1941), pp. 477–516.

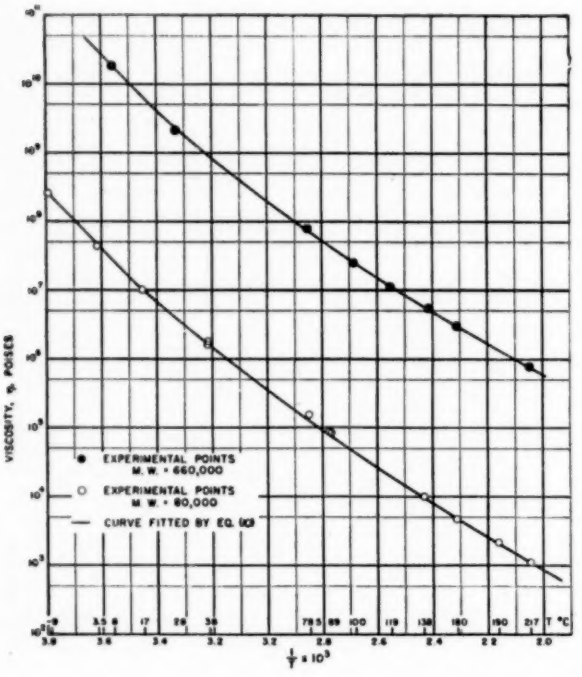


FIG. 2. Comparison of experimental data and Eq. (10) for polyisobutylene of two different molecular weights [data from Fox and Flory (reference 2)].

(d) Finally, closeness of fit was judged by a graphical comparison of theoretical *vs* experimental points as well as by the simple average deviation in $\log\eta$ of the experimental points from the fitted curve without going through a least square procedure.

The constants of Eq. (10) and the average deviations are listed in Table I.

Figure 1 is a typical example for glass No. 10 studied by Robinson and Peterson.[3] The equation evidently fits the melt viscosity data extremely well.

Figure 2 shows the applicability of Eq. (10) to the melt viscosity of a high polymer, polyisobutylene. The data were taken from the work of Fox and Flory.[4]

Equation (10) is also applicable to phenol-formaldehyde novolak resins. Two of Jones[4] curves were fitted with good accuracy. Unfortunately, all the data do not cover a wide enough temperature range or are not sufficiently precise to obtain an unambiguous determination of the constants of Eq. (10). Thus, correlation of viscosity characteristics with molecular weight cannot be carried out without further experimental data.

The data of Spencer and Dillon[1] on polystyrene were not found to be extensive enough to yield an unambiguous determination of the parameters of Eq. (10).

Viscosity data on water, taken from the paper by Kampmeyer,[13] are fitted so closely by Eq. (10) that no deviation is observable on a graph. In this case the deviations were calculated on a viscosity rather than $\log\eta$ basis. The agreement is excellent. It is of interest to note that the value for E is only slightly lower than the value of the over-all activation energy measured at 150°C. Above this temperature, therefore, there is very little contribution to the over-all activation energy resulting from structural changes.[11]

[13] P. M. Kampmeyer, J. Appl. Phys. **23**, 99 (1952).

IV. DISCUSSION

It has been shown in the previous sections that a relation, which takes into account structural changes via the formalism of short range order, can be used to describe the viscosity-temperature properties of materials which are characterized by highly variable overall activation energies for viscous flow. An unambiguous determination of the constants of this equation is only possible when rather accurate viscosity data are available over a wide range of temperature. More experimental data are certainly needed before the general applicability of these ideas can be determined.

It would be extremely interesting to correlate viscosity results with structural studies by means of x-ray and neutron diffraction. It should be possible, by means of diffraction studies, to determine whether short range order as a function of temperature is describable by means of the simple Eq. (7) used here. Further, it should be possible to evaluate U and T_0 directly from such structural studies.

At a more empirical level, it would appear worth while to correlate the constants of Eq. (10), as determined from viscosity measurements, with the composition of mixtures, substitution of other components, molecular weights and molecular weight distributions, etc. For example, phenol-formaldehyde novolak resins could be investigated in some detail with respect to the effect of altering the molecular weight, phenol and water content. Phenol[4] and water[14] are known to be very efficient plasticizers for these resins. With a combination of different techniques of measurement[4,14] an extremely wide range of temperature can be investigated. Similarly, it is well known that the addition of a platicizer has a tremendous influence on the viscosity-temperature characteristics of many high polymers[15] and consequently interesting structural information may be derivable from detailed viscosity-temperature studies. Sufficiently precise data over a wide enough temperature range do not seem to be available at the present time and considerably more experimental work is needed in this field.

[14] G. J. Dienes, J. Colloid Sci. **4**, 257 (1949).

[15] See, for example, W. M. Gearhart and W. D. Kennedy, Ind. Eng. Chem. **41**, 695 (1949); G. J. Dienes and F. D. Dexter, Ind. Eng. Chem. **40**, 2319 (1948).

The Output Signal-to-Noise Ratio of a Power-Law Device

NELSON M. BLACHMAN
U. S. Office of Naval Research, Washington 25, D. C.
(Received January 8, 1953)

An nth law device is fed by a sinusoidal signal and narrow band random noise whose spectrum lies close to the signal frequency. Its output spectrum is different from zero only in the vicinity of harmonics of this frequency. The ratio R' of output signal power to output noise power in the vicinity of the mth harmonic is computed in terms of the input signal-to-noise power ratio R. In general, $R'(R)$ involves a couple of hypergeometric functions, but for small R we have $R'\cong\Gamma^2[(m+n)/2+1]R^m/(m!)^2\Gamma(n+1)$ for $m\neq 0$, $R'\cong\Gamma^2(\frac{1}{2}n+1)/[\Gamma(n+1)-\Gamma^2(\frac{1}{2}n+1)]$ for $m=0$. For large R we have $R'\cong 2R/(m^2+n^2)$.

USING a Fourier representation for random noise, Rice has found[1] the output spectrum of a square-law device fed by a sinusoidally modulated carrier and narrow band random noise. By the use of correlation function methods, he has determined[1] the output spectrum of a linear rectifier, and he[1] and Middleton[2] have shown how the output spectrum may be calculated for the general nonlinear device fed by a sinusoidally modulated carrier and narrow band random noise. Middleton has applied[3] this work to the determination of the low frequency output spectra of power-law devices. Recently Davenport, using Rice's characteristic function technique[1] and integrating over the part of the output spectrum lying near the input signal frequency, has obtained[4] the signal frequency output signal-to-noise ratio for a power-law device fed by an unmodulated carrier and narrow band random noise.

By the use of first-order statistics alone, on the other hand, Rice has found[1] the dc and *total* low frequency output power of some nonlinear devices fed by random noise and one, two, or three sinusoids. It is the aim of the present paper to extend such direct methods to the consideration of the power-law device fed by a sinusoidal signal and narrow band random noise, determining the signal-to-noise ratio for the output components lying in the vicinity of any harmonic of the input signal. It suffices to use first-order statistics because the precise shape of the input noise spectrum cannot affect the result; it influences only the rapidity with which the input amplitude and phase change.[1,2] Such an approach has been used advantageously in the calculation of the (low frequency) output signal-to-noise ratio of a limiter and discriminator fed by a frequency modulated carrier and narrow band random noise.[5]

We shall consider the effect of passing a sinusoidal signal and narrow band random noise lying close to the signal in frequency through a power-law device. The latter will be assumed to have an output voltage au^n when its input voltage u is positive and to have zero output for negative input. Odd symmetry of the characteristic would merely double the odd harmonic output and annihilate the even; even symmetry would double the even harmonic output and annihilate the odd. The exponent n may have any value exceeding -1. We shall be interested in output components whose frequencies lie in the vicinity of the mth harmonic of the signal frequency.

The assumed input is quasi-sinusoidal; its amplitude and phase will change only slightly during any single cycle. During a cycle when the input amplitude is V and its phase is ϕ, the output has an mth harmonic component of amplitude

$$V'(V)=\frac{a}{\pi}\int_{-\pi/2}^{\pi/2}(V\cos\theta)^n\cos m\theta d\theta = (\text{say})ab_m(n)V^n \tag{1}$$

and phase $m\phi$. (For $m=0$, replace π by 2π.) It will not be necessary to evaluate the integral $b_m(n)$, as it will cancel out. If n is an integer, however, b vanishes for m equal to $n+2$, $n+4$, $n+6$, etc. The output signal-to-noise ratio found in this paper is in these cases the limit as n approaches the integral value.

If the amplitude of the input signal is A and its phase angle is zero, and the rms input noise voltage is σ, the probability that the total input amplitude is V and its phase is ϕ is[1]

$$p(V,\phi)d\phi dV=\exp-\frac{A^2+V^2-2AV\cos\phi}{2\sigma^2}\frac{Vd\phi dV}{2\pi\sigma^2}. \tag{2}$$

The total mth harmonic output power of the device is half the mean square value of V':

$$\tfrac{1}{2}\int_0^\infty\int_0^{2\pi}V'^2(V)p(V,\phi)d\phi dV = \frac{a^2b_m^2(n)}{2\sigma^2}\int_0^\infty V^{2n+1}\exp-\frac{A^2+V^2}{2\sigma^2}I_0\left(\frac{AV}{\sigma^2}\right)dV = 2^{n-1}a^2b_m^2(n)\Gamma(n+1)\sigma^{2n}{}_1F_1(-n;1;-A^2/2\sigma^2). \tag{3}$$

[1] S. O. Rice, Bell System Tech. J. **24**, 46 (1945).
[2] D. Middleton, Quart. Appl. Math. **5**, 445 (1948).
[3] D. Middleton, Proc. Inst. Radio Engrs. **36**, 1467 (1948).
[4] W. B. Davenport, J. Appl. Phys. **24**, 720 (1953).
[5] N. M. Blachman, J. Appl. Phys. **20**, 38 (1949). In another paper, N. M. Blachman, J. Appl. Phys. **20**, 976 (1949), Rice's characteristic function technique is applied to the determination of the spectrum of the discriminator output in the absence of limiting. For the complete solution of the fm problem through the use of correlation methods, see D. Middleton, Quart. Appl. Math. **7**, 129 (1949); **8**, 59 (1950).

(This expression should be doubled for $m=0$, since no $\frac{1}{2}$ is then needed.)

The component of the output in phase with the mth harmonic of the input signal has, at any instant, amplitude $V'\cos m\phi$. Its average value is the amplitude of the mth harmonic output signal component:

$$\int_0^\infty\int_0^{2\pi} V'(V)\cos m\phi\, p(V,\phi)\,d\phi\,dV=\frac{ab_m(n)}{\sigma^2}\int_0^\infty V^{n+1}\exp-\frac{A^2+V^2}{2\sigma^2}I_m\left(\frac{AV}{\sigma^2}\right)dV$$

$$=\frac{ab_m(n)}{m!}\Gamma\left(\frac{m+n}{2}+1\right)A^m(\sqrt{2}\sigma)^{n-m}{}_1F_1\left(\frac{m-n}{2};m+1;-\frac{A^2}{2\sigma^2}\right). \quad (4)$$

The mth harmonic signal-power content of the output is given by half (except for $m=0$) the square of (4); the remainder of the mth harmonic output power, the difference between (3) and (half) the square of (4), represents the mth harmonic noise output power. The mth harmonic output signal-to-noise power ratio is therefore

$$R'(R)=\frac{\Gamma^2\left(\frac{m+n}{2}+1\right)R^m\,{}_1F_1^2\left(\frac{m-n}{2};m+1;-R\right)}{(m!)^2\Gamma(n+1)\,{}_1F_1(-n;1;-R)-\Gamma^2\left(\frac{m+n}{2}+1\right)R^m\,{}_1F_1^2\left(\frac{m-n}{2};m+1;-R\right)}, \quad (5)$$

where $R=A^2/2\sigma^2$ is the input signal-to-noise power ratio.

For large R, (5) is approximately

$$R'(R)\cong 2R/(m^2+n^2)\ (R\gg 1). \quad (6)$$

For small R, (5) is approximately

$$R'(R)\cong\left.\begin{aligned}&\Gamma^2\left(\frac{m+n}{2}+1\right)R^m\Big/\\ &\qquad (m!)^2\Gamma(n+1)\ (R\ll 1,\quad m>0),\\ &\cong\Gamma^2(\tfrac{1}{2}n+1)/[\Gamma(n+1)(1-\tfrac{1}{8}n^2R^2)\\ &\qquad -\Gamma^2(\tfrac{1}{2}n+1)]\ (R\ll 1,\quad m=0)\end{aligned}\right\}. \quad (7)$$

With a large input signal-to-noise ratio, the output signal-to-noise ratio is seen to be larger the smaller are m and $|n|$. Indeed, when $m=n=0$, R' is infinite, since the output of an ideal limiter (the power-law device becomes an ideal limiter when n is zero) fed by a high frequency input can contain no low frequency fluctuations.

When R is small, R' is again larger the smaller is m. However, for a given m, R' is now largest when n equals m. The high output signal-to-noise ratio at low frequencies ($m=0$) is perhaps deceptive. It is due to the fact that a "signal" output (rectified noise) is obtained even in the absence of an input signal; this dc output component is regarded as part of the output signal for the sake of consistency in terminology.

For integral values of n, ${}_1F_1(-n;1;-R)$ is a polynomial of degree n in R, while ${}_1F_1((m-n)/2;m+1;-R)$ is a polynomial in R, $1/R$, and e^{-R} when $m-n$ is even and is expressible in terms of the modified Bessel functions I_0 and I_1 when $m-n$ is odd. For example, the signal frequency ($m=1$) output signal-to-noise ratio of an ideal limiter ($n=0$) is

$$R'(R)=\frac{\pi R[I_0(\frac{1}{2}R)-I_1(\frac{1}{2}R)]^2}{4e^R-\pi R[I_0(\frac{1}{2}R)-I_1(\frac{1}{2}R)]^2}. \quad (8)$$

Using (6) and (7), we find that for small R, (8) is approximately $R'(R)\cong\frac{1}{4}\pi R$, while for large R it is approximately $R'(R)\cong 2R$. For a linear device ($n=1$), we find from (5) that in the region of the signal frequency ($m=1$) R' is identically equal to R, as it clearly must be. Putting m equal to 0 and n equal to 1 or 2 in (5), we obtain low frequency signal-to-noise ratios for the linear and quadratic rectifier which accord with Rice's values.[1] Setting m equal to 1, we obtain values for the signal frequency signal-to-noise ratio which agree with Davenport's somewhat more complicated expression[4] at least for extreme values of R with n equal to 0, $\frac{1}{2}$, or 1.

The total power output of the device—signal and noise—is the summation of (3) from $m=0$ to $m=\infty$. It is more easily determined, however, by noting that the total power output is the mean value of

$$a^2\int_{-\pi/2}^{\pi/2}(V\cos\theta)^{2n}d\theta/2\pi=a^2b_0(2n)V^{2n}. \quad (9)$$

Thus, the total power output is obtained by substituting $2b_0(2n)=\pi^{-\frac{1}{2}}\Gamma(n+\frac{1}{2})/\Gamma(n+1)$ for $b_m{}^2(n)$ in (3). Notice that as R varies, the fraction of the total power appearing in each harmonic region remains constant. While the power in any harmonic region is finite and equal to (3) for $n>-1$, the *total* power is finite and equal to (9) only if $n>-\frac{1}{2}$. The total *signal* output is finite for $n>-1$.

Likewise, the total signal output wave is best ex-

pressed not as a Fourier summation of (4) but as the average instantaneous output corresponding to the instantaneous signal input. Thus, if the latter is $A\cos\omega t$, the total signal output wave is

$$\frac{a}{(2\pi)^{\frac{1}{2}}\sigma}\int_0^\infty u^n \exp-\frac{(u-A\cos\omega t)^2}{2\sigma^2}du, \qquad (10)$$

since the input contains a Gaussian noise of rms value σ. While (10) is in general difficult to put into more convenient form, it can be expressed in terms of the error function when n is an integer. For example, when $n=0$, the signal output is $a\,\mathrm{erf}((2R)^{\frac{1}{2}}\cos\omega t)$; for $n=1$ it is $a\sigma[(2\pi)^{-\frac{1}{2}}\exp(-R\cos^2\omega t)+(2R)^{\frac{1}{2}}\cos\omega t\,\mathrm{erf}((2R)^{\frac{1}{2}}\cos\omega t)]$; and for $n=2$ it is $a\sigma^2[(R/\pi)^{\frac{1}{2}}\cos\omega t\exp(-R\cos^2\omega t)+(1+2R\cos^2\omega t)\,\mathrm{erf}((2R)^{\frac{1}{2}}\cos\omega t)]$. However, it does not appear possible to express the total signal *power* in convenient form even for integral n. We can nevertheless observe that when the signal-to-noise ratio is large, the signal comes through as if the noise were not present, while when the signal-to-noise ratio is small, as we see from (7), the principal component of the output signal is the dc term, the amplitudes of the higher harmonics being of successively smaller orders of magnitude.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 6 JUNE, 1953

The Electromagnetic Transmission Characteristics of the Two-Dimensional Lattice Medium*

HERBERT S. BENNETT
Dynamic Electronics—New York, Inc., New York, New York

(Received July 29, 1952)

This paper investigates the behavior of the two-dimensional lattice medium in the presence of electromagnetic energy. The conditions for transmission of energy by the medium are derived and studied. In particular, the effects of the various parameters of the problem (e.g., obstacle size, obstacle shape, obstacle spacing, polarization of the electromagnetic energy) on the average propagation constant and impedance function of the medium are considered. The Wigner-Seitz cellular method is applied to the solution of the boundary value problem involving a lattice of elliptic cylindrical obstacles. Various approximate solutions are studied and appropriate functions formulated. Calculations are shown using some of these functions. Practical applications of the theoretical results are indicated, and suggestions for further research are enumerated.

SECTION I

Introduction

1. Purpose

THE purpose of this dissertation is to investigate the behavior of the two-dimensional lattice medium or geometry in the presence of electromagnetic energy. In particular, its behavior will be studied when this energy is allowed to flow in some desired direction. In other words, those characteristics of the lattice geometry which are of particular importance to the transmission of energy through the structure will be critically examined and correlated with the field impedance concept.

Starting from basic principles, the effects of the various parameters entering into the problem will be studied in order to determine how each—individually and in combination—affects the transmission characteristics. The results obtained from this approach should yield firmer design criteria for the various applications of lattice structures. The unification and synthesization of over-all effects will permit a more accurate theoretical prediction of behavior and thus reduce the amount of trial and error experimentation required to achieve practical designs.

2. Definitions

While a complete glossary of terms is contained in Appendix A, it would be desirable at this point to elaborate on some of the terms used in the preceding paragraphs. The term "lattice geometry or medium" is used to signify a regular or coherent arrangement of physical obstacles. This regularity causes a repetition of structure to occur as one proceeds outward along a straight line from an arbitrary origin in multiples of a certain base period or line length appropriate for the origin and direction in question.

We shall choose a rectangular lattice, where, in the plane perpendicular to the z axis, the two primary directions are perpendicular. The periodicity in each of these two directions is such as to yield a rectangular unit cell which contains all the required geometric variations. These cells may be placed down side by side to cover the entire plane and form a lattice medium with the periodicities specified by the two perpendicular sides of the rectangular lattice. The rectangular unit cell is illustrated in Fig. 1 using an obstacle of elliptic cross section.

The term "two-dimensional" is not intended to imply

* This paper is based on research performed at the Polytechnic Institute of Brooklyn, New York, in partial fulfillment of the requirements for the degree of Doctor of Electrical Engineering.

that the obstacles are two-dimensional but rather that the lattice periodicity is limited to two dimensions. In the third dimension the structure is homogeneous. Thus as one proceeds along the z axis at any given x and y value there is no change in properties of the medium. The cross-sectional picture obtained by passing a plane perpendicular to the z axis of the infinite lattice medium or geometry at an arbitrary value of z is identical with that obtained for every other value of z from $+\infty$ to $-\infty$. (In practical applications this uniformity need only extend sufficiently in the z direction to make it large compared to the cross-sectional dimensions of the unit cell and also large compared to λ, the free-space wavelength of the electromagnetic radiation.)

The term "geometry or medium" as distinguished from "structure" is intended to imply periodicity from $x=-\infty$ to $x=+\infty$ and from $y=-\infty$ to $y=+\infty$ and extension from $z=-\infty$ to $z=+\infty$. The term "structure" is intended to imply a lattice of finite extent in at least the y dimension (the direction of propagation). Thus the lattice structure might extend from $y=-y_1$, to $y=\infty$ but with double infinities in both the x and z directions.

Finally, the term "transmission characteristics" is intended to signify the key quantities which measure the average progression of electromagnetic energy through a lattice structure. We shall see later that these are the propagation constant (or index of refraction) and the impedance property representing the discontinuity encountered by electromagnetic energy passing from a uniform homogeneous medium into the lattice medium. The very name "transmission characteristic" is intended to imply that the cases of principal interest are those where the particular frequency or frequencies in question are not prohibited from propagating through the lattice medium. (Of course, by studying all the transmission ranges one automatically studies and delineates the boundaries of the nonpropagating regions in the frequency domain.)

3. *Elliptic System*

For quantitative analysis, the shape of the obstacle within the unit cell is limited by the number of two-dimensional orthogonal coordinate systems in which at least the scalar wave equation is separable and soluble. There are only eleven distinct coordinate systems in which the three-dimensional wave equation is separable.[60]†

In our so-called two-dimensional problem it was apparent that the elliptic cylindrical coordinate system offered the greatest flexibility in so far as obstacle shape in the x, y plane was concerned. The elliptic cylinder has as its limiting cases, the infinitely thin strip and the circular cylinder. Intermediate between these limits the variety of relations available between the major and minor axes allow almost any two-dimensional obstacle to be represented with fair approximation by an elliptic cross section.

4. *Scope of Investigation*

In order to accomplish the purpose of the investigation, our two-dimensional lattice medium problem will be approached from both the classical "cavity" method and the, by now no less classical, network or engineering approach employing the impedance concept.

In the former approach[44] the complete solution of Maxwell's equations for a given electromagnetic field problem is intimately connected with the particular geometry of the system. After setting up the differential equations of the field in the appropriate coordinate system, each problem is treated as a peculiar and individual entity. The solution of the equations, with appropriate boundary conditions, for the region under consideration is the end objective.

In the latter approach, on the other hand, the emphasis is on setting up general classes of solutions. These solutions are tied in with electrical circuit theory in general and iterative circuit theory (e.g., filter networks and transmission lines) in particular. The "cavity" solution is recast into such a form that it represents an equivalent network. Associated with this network is a certain propagation function and a certain characteristic impedance function. This approach implies that a certain direction of primary energy transmission is of particular interest. We shall find that in our case this is so. The microwave network approach masks the "microscopic" solution of the E and H fields for the lattice structure and emphasizes the "macroscopic" behavior of the structure as a transmitter of electromagnetic energy. In fact it presents these characteristics of the lattice structure in such a form as to facilitate the calculation of the performance of a larger over-all transmission system of which the lattice structure is merely one component. This latter facility is perhaps the most important single contribution of the "engineering" approach to the practical application of microwave structures in complicated electromagnetic systems.

We will study a medium which is periodic in the direction of propagation as well as in the transverse direction.[5,10] This is the case of the infinite lattice geometry or medium. A physical explanation of wave propagation in both lattice geometries and lattice structures will be furnished. Unifying impedance concepts will be established tying together the various contributions already made in this field.

In Sec. III a mathematical solution of the two-dimensional lattice medium composed of elliptic cylinders will be carried out applying the Bloch wave solutions of the Wigner-Seitz cellular theory of solids[48] to a much more general situation than has heretofore been accomplished. This will include the derivation of the basic functions to allow the calculation of both the

† The superscript number will hereafter refer to a specific numbered reference in the Bibliography, Appendix B.

propagation and characteristic impedance functions of the medium. Calculated results will be shown.

Section IV will contain some general conclusions as well as discussion of the practical applications of the foregoing analyses and some problems considered worthy of further research.

5. *State of the Art*

To conclude this introductory section we will give a brief qualitative bird's-eye view of work already accomplished in this field.

In the work done to date on the lattice medium the tendency has been to compartmentalize rather than unify the results. Thus we find that the delay type of lattice medium (the so-called artificial dielectrics) has been considered separately from the "speed-up" type (the metal plate lenses). Also, almost universally the quasi-static case (λ very much larger than any of the dimensions of the unit cell) has been the only one considered. While it is true that this case has some very important practical applications, it should not be overlooked (see Sec. IV) that a vast unexplored field of application exists for lattice media operated at other than the simplest conditions. It was because of the lack of generalized analyses of the lattice media unifying the consideration of the effects of the various significant parameters that the investigation upon which this paper is based was undertaken.

To date, the analysis of the lattice medium has been confined principally to two major approaches: (a) analogy with the molecular theory of matter, (b) transmission line filter analogy.

In considering the lattice medium as a large scale model of a homogeneous dielectric ("a" above) two approaches have been used.

(1) Kock[36] considered that λ was so large that the geometry of the particle was of secondary importance. The particle obstacle acquires a dipole moment under the influence of an impressed electric field. Each particle is completely independent of the others and thus the polarization of the gross lattice medium is obtained by a simple summation. Not only are the particles required to be very small with respect to wavelength, but their spacing must be very large compared to the particle size. This is essentially an analogy to the molecular theory of gases.

Corrections of the above to take into account particles of larger size compared to spacing have been worked out. The Clausius-Mossotti formula is one such attempt. Corkum[22] and Estrin[26] also have presented a correction for the "bulking" effect of closely packed spheres. None of these corrections, however, remove the restrictions that both particle size and spacing must be small compared to the free-space wavelength of the incident electromagnetic energy.

(2) Lewin,[39] Pasternack,[43] and Slepian[51] have extended the above approach by analogy with the molecular theory of liquids and solids. Here interaction between particles is taken into account. This allows a start on an analysis which does not require λ very large compared to spacing and obstacle size. However, only Slepian removed this restriction to any significant extent in a quantitative way. But Slepian only considers the circular cylindrical type of obstacle, and the unit cell boundary conditions are matched at the absolute minimum of points.

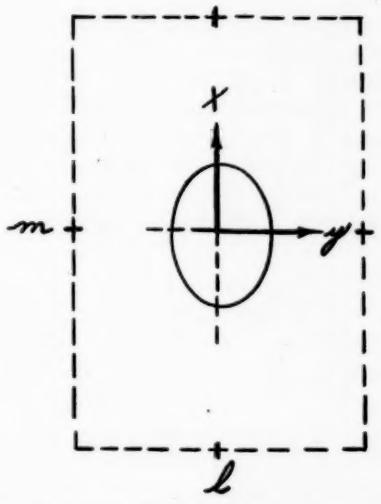


FIG. 1. Illustration of rectangular unit cell with an obstacle of elliptic cross section.

The method "*b*" above—which uses the Wigner-Seitz cellular concept coupled with Bloch wave solutions[5]—requires that the wave equation be satisfied throughout the lattice medium but so "forces" the mode solution appropriate for scattering by the obstacle in the cell as to make the solution of propagation through the lattice medium be represented by Bloch waves. Thus the requirement for satisfying the wave equation over all of space is reduced to satisfying it within and on the boundaries of the unit cell of Fig. 1.

The transmission line-filter analogy set forth by Cohn,[20] Brown,[13] and Jones[15] considers that the lattice medium is made up of a stack of parallel plate wave guides; each wave guide being loaded at periodic intervals along its length by the lattice obstacle. This method in so far as it has been carried to date does not take into account the interaction between obstacles in the direction of propagation. Neither does it take into account, in a quantitative way, the case when λ is smaller than the spacing of the obstacles in the direction of propagation. By making use of work on the impedance properties of infinite plane gratings by MacFarlane (quasi-stationery, λ greater than transverse spacing) and by Marcuvitz (application of Schwinger's integral equation techniques—no restriction on λ in relation to *transverse* spacing in the most general application), accurate representations of the network parameters of transversely infinite plane gratings may be obtained. However, to date the addition of an infinite number of

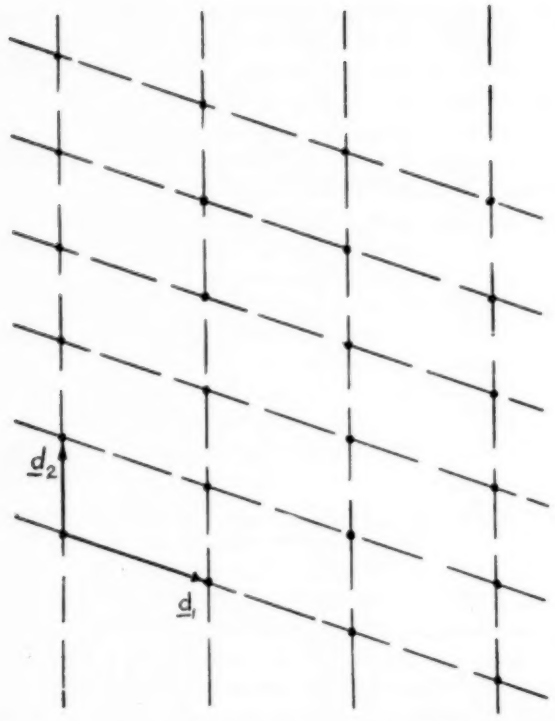


FIG. 2. Cross section of a lattice medium which is periodic in two directions (axis of cylinders are perpendicular to the plane of the drawing).

these gratings to form a lattice medium has been accomplished only for λ greater than spacing in direction of propagation and for interaction in the direction of propagation between nearest neighbors only. Brillouin[10] indicates a formal solution for interaction between more than nearest neighbors, but it has not been applied to quantitative solutions of lattice mediums to date; neither has the Schelkunoff field impedance concept been applied to the lattice structure previous to this dissertation.

SECTION II

Lattice Structures—General Theoretical Considerations

1. Cell Concept

In this paper we limit ourselves to media with periodicity in only two directions (see Fig. 2). The loss in generality is not great since all the basic elements of the problem show up in two dimensions just as well as in three but with a simplification of notation and computation. Furthermore, there is evidence that the inherently mechanically self-supporting cylindrical type of obstacle is finding a constantly increasing application in practical engineering designs.

The lattice medium has a periodicity in the direction of propagation from $+\infty$ to $-\infty$ in addition to the transverse infinite periodicity. In Fig. 2 if we take $\mathbf{d}_1$ and $\mathbf{d}_2$ as the two perpendicular basis vectors of the rectangular unit cell, the vector coordinate of any point in the lattice from the particle or point chosen as the origin is

$$\mathbf{r}_{l_1 l_2}=l_1\mathbf{d}_1+l_2\mathbf{d}_2, \tag{2.1}$$

where l_1 and l_2 are integers varying between $+$ and $-\infty$.

Whereas in the case of a planar grating or in the case of a single obstacle case, we adjusted the form and constants of our mode solutions to give results which did not go to infinity as we proceeded large distances from the obstacles, the second periodicity in the two-dimensional lattice medium "forces" the repetition of the boundary conditions at each cell boundary. For the moment let us deal with radiation propagating along the y axis in Fig. 2. Again by symmetry considerations we find that the wave must continue to propagate in the y direction. This forms the basis for the Wigner-Zeitz[57,58] method of treating the theory of solids. The general solution for a wave proceeding through the lattice geometry can be given in the form

$$\pi_z=\psi(\mathbf{r}, k)e^{-i(\omega t-K_T y)}, \tag{2.2}$$

where ψ is the amplitude factor, $e^{-i(\omega t-K_T y)}$ is the phase factor, k is the free-space propagation constant, and K_T is the average or over-all lattice propagation constant corresponding to a particular value of k.

Equation (2.2) arbitrarily defines a set of so-called Bloch waves. These waves must meet all the electromagnetic boundary conditions imposed by the boundary of the perfectly conducting obstacle and in addition must meet the requirements that corresponding points on the cell boundary side faces (across which no energy flows) have equal values of π_z, whereas corresponding points on end faces retain equality in $\psi(\mathbf{r}, k)$ but have a phase shift of e^{iKl} in π_z. Because of these conditions K may be considered as the equivalent propagation constant (though as we shall see later in Sec. III, K may be a very complicated function of k) through the lattice medium.

Thus we have reduced the problem of the lattice medium (infinite and periodic in 2 directions, infinite and constant in the third direction) to the solution of a two-dimensional wave equation within and on the boundaries of the unit cell.

$$\nabla^2\pi_z+k^2\pi_z=0. \tag{2.3}$$

Equation (2.3) must hold all over the cell except inside the perfectly conducting obstacles and must meet the normal boundary conditions on a perfect conductor.

On the cell boundaries (end faces), where "$\mathbf{a}_1$" serves to translate an end-face point to a corresponding point one cell away, we have

$$\begin{aligned}\psi(\mathbf{r}_1, k)&=e^{iKl}\psi([\mathbf{r}_1+\mathbf{a}_1], k),\\ \bar{n}\cdot[\nabla\psi(\mathbf{r}_1, k)]&=e^{iKl}\bar{n}\cdot[\nabla\psi([\mathbf{r}_1+\mathbf{a}_1], k)];\end{aligned} \tag{2.4}$$

$\bar{n}$ is normal to the cell face in question. For side faces the condition is the same except the e^{iKl} factor is not present.

For H plane polarization (i.e., the case where only the E_z component of the total vector E field differs from zero and $H_z=0$ in general throughout the cell) we have at the obstacle boundary

$$\pi_z=0, \tag{2.5}$$

and for E plane polarization (i.e., $H_z\neq 0$, $H_x=H_y=E_z=0$, $E_x\neq 0$, $E_y\neq 0$)

$$\partial\pi_Z^*/\partial n'=0. \tag{2.6}$$

(The asterisk denotes the E plane case.) n' is normal to the obstacle surface at any point.

2. *Propagation Through Infinite Lattice Media*

Let us discuss now the two primary factors required to describe the electromagnetic transmission behavior of lattice structures or media, namely, index of refraction (or the propagation constant) and the lattice junction or impedance to power transmission characteristics.

According to our terminology the term "structure" implies that the lattice medium does not extend out to $-\infty$ in the direction of propagation. (It also might not extend out to $+\infty$ in the direction of propagation). The significance of mentioning both the terms "media" and "structures" in this section is tied in with the manner of calculating the two primary factors mentioned above.

It should be emphasized again at this point that our principal efforts are oriented toward analyzing lattice structures under conditions conductive to power transfer by Bloch waves. Information gained concerning the ranges of nontransmission will be only a by-product of the basic effort.

Now let us consider the determination of the propagation constant for the transmission of electromagnetic energy through an infinite lattice medium. The basic philosophy behind the solution of the transmission of electromagnetic energy through a lattice medium concerns the symmetries caused by the coherent repetition of the obstacles in any plane section perpendicular to the infinite longitudinal axis of the obstacle. We shall so arrange our Cartesian coordinate axes as to have the z axis coincide with the infinite longitudinal axis of any one of the cylindrical obstacles (an arbitrary cross section is assumed for the cylindrical obstacle). The plane section is taken at $z=0$ and the $x-y$ axes so oriented that the smallest cell in the $x-y$ plane whose infinite repetition from $x=-\infty$ to $x=+\infty$ and $y=-\infty$ to $y=+\infty$ will form the lattice medium is a rectangle (see Fig. 3). The symmetries are such that if a plane wave exists in the medium with its direction of propagation along the positive y axis, at each successive cell entrance face the planarity is maintained. In fact, whatever the amplitude distribution along the entrance face is, it is repeated at the next cell entrance face. The only difference in the character of the electromagnetic energy which we observe, if we restrict ourselves to sampling at cell entrance faces only, is that the phase of the constant-phase front is changed by a certain amount $\pm\delta$ from what it was at the preceding cell face. It is this coherent phase change at successive entrance faces, while maintaining the direction of propagation and transverse amplitude distribution at each cell entrance face, which allows the complicated infinite boundary value problem to be solved by Bloch waves satisfying boundary conditions on the unit cell boundaries and on the surface of the perfectly conducting obstacle within the cell.

It should be noted that the above outlined repetition of conditions at successive entrance faces is very basic to the assumption of our particular cell boundary conditions. To what may we attribute this repetition? There are two possible explanations:

(1) The obstacle at the center of the cell is so small that the cell entrance and exit faces are so many wavelengths from the obstacle that only the principal or exciting mode is propagating all the transmitted energy and there are no interactions between cells in the y direction. This is the case considered by S. B. Cohn.[20] In this case it is possible to superimpose the effects of successive planar gratings to form an iterative structure of uniform transmission line sections shunted by identical four-poles. This approach takes into account the detailed geometric shape of a single obstacle and also (except in its simplest form) takes into account large cross-sectional obstacle size compared to transverse (x direction) obstacle spacing.

(2) A second, more general explanation of the reason for the preservation of transverse amplitude distribution at cell entrance faces takes into account the infinite interactions between any single obstacle and all others in the $y=+\infty$ and $y=-\infty$ directions. By taking these interactions into account we remove the restriction that the obstacle size must be small compared to the cell size in the y direction. In this general case we may visualize the behavior of the total E and H fields as the energy

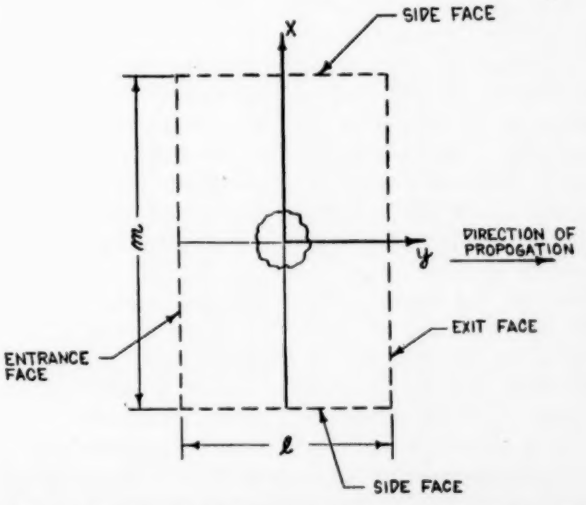


FIG. 3. Rectangular cell representation for a lattice medium composed of cylindrical obstacles of arbitrary cross section disposed in a rectangular array.

proceeds through the unit cell. Let us assume that at the entrance face there exists a plane constant phase front and a certain arbitrary amplitude distribution along x for both the total E and total H fields.

As the energy penetrates the cell the plane phase front is distorted, since the total fields are the vector sum of the incident and diffracted fields and the diffracted fields from cylindrical obstacles are not plane except at great distances from the obstacle. Thus, as we go from the entrance face to the mid-point of the cell, the distortion of the total E and H fields steadily increases and this has the effect of changing the energy transmission from a plane phase front to some more complicated shape. There are components of energy transmission in other than the y direction. These are not really energy transmitting modes since they damp out before reaching the cell side faces. Therefore, the effect of the infinite interactions is to force the cell boundaries to be the "far-field" region of the cell.

As the energy field proceeds from the midpoint of the cell to the cell exit face, a restoration effect starts which results in a plane-phase front (perpendicular to the y axis) again and restores the original amplitude distribution of the total E and H fields along X. If we sampled the total E and H fields only at the entrance and exit faces we would find that the energy propagation phenomenon could be specified by a wave with a plane constant-phase front perpendicular to the Y axis and an amplitude which is a certain function of X. These are known as Bloch waves. The average lattice propagation constant in the Y direction is either greater or less than the free-space propagation constant k_T depending on the relationship of obstacle geometry to electromagnetic field polarization. (We are assuming, of course, that our perfectly conducting lattice obstacles are in free space. This is the practical situation; but any other perfect dielectric medium would not introduce any procedural changes in the analysis which follows.)

We will now formulate the problem analytically. We choose the Cartesian coordinate system for the portion of the analysis which does not consider the fields in the interior of the cell in detail, because the rectangular shape of the unit cell of the lattice medium lends itself naturally to such a system. The total vector fields must satisfy Maxwell's equations throughout the free-space matrix of the lattice medium and, in addition, satisfy the required boundary conditions on the perfectly conducting obstacle surface and on the cell boundaries. (See Eqs. 2.4 through 2.6.) We also take advantage of the fact that for waves whose phase fronts remain parallel to the z axis, we can define any arbitrary field as the linear sum of an E plane and an H plane partial field as defined in subsection 1 above. (We shall also see in the next section that this division into partial fields is especially appropriate for analysis of our lattice medium since the medium reacts differently to the E and H plane polarizations.) The linear nature of the electromagnetic fields in our lattice medium allows us to decompose an arbitrarily polarized field into its component E plane and H plane parts, solve each part separately, and add the solutions to secure the total solution that is due to the total field.

The Hertz vector concept coupled with the preceding affords us a very convenient analytical tool for solving our problem. Both the E plane and H plane fields are completely specified by their respective Hertz components along the z axis, π_z and π_z^*.

Remember that

$$\bar{E}=\nabla x(\nabla x\bar{\pi})=\nabla(\nabla\cdot\bar{\pi})-\nabla^2\bar{\pi},$$
$$\bar{H}=\epsilon_0\frac{\partial}{\partial t}(\nabla x\bar{\pi}), \tag{2.7}$$

and similarly that

$$\bar{H}^*=\nabla x(\nabla x\bar{\pi}^*) \quad \text{and} \quad \bar{E}^*=-\mu_0\frac{\partial}{\partial t}\nabla x\bar{\pi}^*. \tag{2.8}$$

The relationships between the Hertz vectors and the **E** and **H** vectors in rectangular coordinates are for the E plane case, assuming the following sinusoidal time variation:

$$E_x^*=-\mu_0\frac{\partial}{\partial t}\left(\frac{\partial}{\partial y}\pi_z^*\right)=i\omega\mu_0\frac{\partial\pi_z^*}{\partial y}$$
$$=ik\left(\frac{\mu_0}{E_0}\right)^{\frac{1}{2}}\frac{\partial\pi_z^*}{\partial y}, \tag{2.9}$$
$$E_y^*=\mu_0\frac{\partial}{\partial t}\left(\frac{\partial}{\partial x}\pi_z^*\right)=-ik\left(\frac{\mu_0}{E_0}\right)^{\frac{1}{2}}\frac{\partial\pi_z^*}{\partial x},$$
$$H_z^*=k^2\pi_z^*.$$

Similarly for the H plane case,

$$E_Z=k^2\pi_z,$$
$$H_x=\epsilon_0\frac{\partial}{\partial t}\left(\frac{\partial\pi_z}{y}\right)=-i\omega\epsilon_0\frac{\partial\pi_z}{\partial y}=-ik\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}\frac{\partial\pi_z}{\partial y}, \tag{2.10}$$
$$H_y=-\epsilon_0\frac{\partial}{\partial t}\left(\frac{\partial\pi_z}{\partial x}\right)=i\omega\epsilon_0\frac{\partial\pi_z}{\partial x}=ik\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}\frac{\partial\pi_z}{\partial x}.$$

Remembering that $k=$ free-space wave number $=2\pi/\lambda_0$, we see that the scalar wave equation is appropriate for E_z in the H plane case and for H_z^* in the E plane case. Therefore, since k is a constant in so far as the x, y, z coordinates are concerned, and remembering that $E_z=k^2\pi_z$ and $H_z^*=k^2\pi_z^*$ we have

$$\nabla^2\pi_z+k^2\pi_z=0,$$
and
$$\nabla^2\pi_z^*+k^2\pi_z^*=0. \tag{2.11}$$

Equations (2.11) must be satisfied at all points in the free space in the lattice medium and also on the surface

of the perfectly conducting obstacles. As we have seen, because of the infinite repetition of unit cells, the above requirements for Eq. (2.11) are retranslated into boundary conditions Eq. (2.4) through (2.6) for the unit cell.

Now we must relate the solution in terms of the set of Bloch waves Eq. (2.2) with the boundary requirements of Eq. (2.11).

The first step is to assure that ψ of Eq. (2.2) satisfies the electromagnetic boundary conditions on the surface of the obstacle. The mode solutions are very convenient to use, provided the shape of the object is amenable to treatment in one of the separable coordinate systems. Thus,

$$\psi=\sum_{m=0}^{n} A_m F_1(\epsilon, k)F_2(\eta, k), \tag{2.12}$$

where ϵ and η are the orthogonal coordinates of the coordinate system which for $\epsilon=a=$ constant specifies the surface of the obstacle. Both ϵ and η are functions of x and y. Thus, from Eqs. (2.5) and (2.6) we so arrange F_1 that

$$\begin{aligned}&[F_1(\epsilon)]_{\epsilon=a}=0,\\ \text{and}\qquad &\\ &[\partial F_1{}^*/\partial\eta']_{\epsilon=a}=0.\end{aligned}\tag{2.13}$$

The constants A_m are evaluated from Eq. (2.4). The number n of these that may be evaluated depends on the number of pairs of points at which Eq. (2.4) is evaluated. Theoretically n should be infinite. In practice the points have been restricted to a pair at $y=0$, $x=\pm l/2$ and $x=0$, $y=\pm m/2$. This is probably far from adequate when the obstacle occupies an appreciable proportion of the volume of the unit cell. In Sec. III we will discuss this point further.

In evaluating the constants A_m with Eq. (2.4), one gets a set of algebraic equations where each equation individually is equal to zero and there are as many equations as there are unknown constants. In order for this set of equations to yield nontrivial solutions the determinant must equal zero. This has the effect of determining all the constants in terms of one of the set—say A_0, so that

$$\psi=A_0\sum_{m=1}^{n} F_1(\epsilon, k_T)F_2(\eta, k_T). \tag{2.14}$$

Because of the form of Eq. (2.4) the average lattice propagation constant K enters the above-mentioned determinant only in the form $\cot^2(Kl/2)$, if boundary conditions are matched only at the 2 pairs of points $y=0$, $x=\pm l/2$ and $x=0$, $y=\pm m/2$.

Thus we get a transcendental equation relating K and k

$$\cot^2[K_T(l/2)]=-F_3(k_T). \tag{2.15}$$

The "T" denotes those ranges of k_T for which F_3 is negative in order that K_T be real and in order that propagating Bloch waves exist. The functional form of F_3 depends on the form of the cross section of the obstacle, e.g., for circular cylinders F_3 is made up of Bessel functions containing k_T in their argument while for elliptic cylinders F_3 is made up of Mathieu functions.

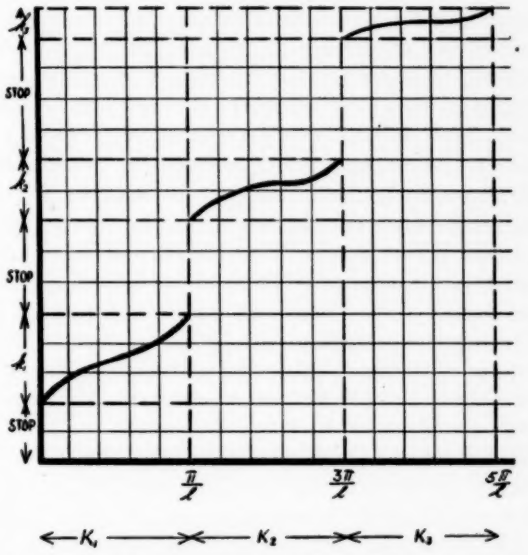


FIG. 4. Pictorial representation of transmission band structure of an infinite lattice medium where k_n *vs* K_n has been separated into single-valued cycles.

It can be seen that in solutions of Eq. (2.15) for any given k there are a doubly infinite number of positive or negative Bloch waves given by

$$\pm\left(K_1+\frac{n2\pi}{l}\right)$$

and

$$\pm\left(\left[\frac{2\pi}{l}-K_1\right]+\frac{n2\pi}{l}\right),$$

where

$$K_1 l/2\leq\pi/2,$$

and where n is any integer from zero to infinity. Similarly, for any given value K_1 there are cycles of k_T for which F_3 is negative. Also at the points $-F_3(k_T)=0$ (corresponding to $K=(2n+1)\pi/l$) a region in k starts for which F_3 is not negative. This portion of the cycle results in specifying the stop or forbidden bands of frequency where energy will not be propagated through a given infinite lattice medium. Depending on whether a pole or a zero occurs first in the K *vs* k characteristic we have a pass or stop band in the low frequency region.

For analysis of energy propagation in infinite lattice media it is very useful to so separate the cycles of K and k so that for the first cycle of K (i.e., K_1) only one single valued band of frequencies k_1 exist. Therefore K_T and k_T specify the first, second, third, etc., *pass* bands where k_T is a single valued function of K_T. This is illustrated in Fig. 4. This separation corresponds to the physical situation in an infinite lattice. Electromagnetic energy propagating in the infinite lattice in the y direction with a fixed frequency or free-space wave number k_T

will have an average lattice propagation constant K_T provided the transverse cell spacing m is less than λ. When m is greater than λ, more than one Bloch wave transmission mode propagates the energy through the lattice. This will be more fully discussed below. (It should be noted that energy propagating in a direction other than along the y axis may be divided into two components—one propagating along the y axis and the other along the x axis. Thus in Sec. III the lattice propagation constants are derived for both the x and y directions.)

The physical situation in a finite lattice structure is not quite as simple. Energy k_T incident from free space on the first face of the lattice structure will tend to excite an infinity of K_T's and depending on the interrelation of many parameters (especially the relationship of the cell dimension "m" to λ), a given number of the K_T's corresponding to several successive cycles [where $K_n = K_{n-1} + (2\pi/l)$] will actually propagate the energy through the lattice. This situation and others concerning the finite lattice structure will be taken up below where the impedance concept will be discussed.

Before proceeding to the discussion of the impedance concept it would be well to summarize the parameters involved in the propagation of electromagnetic energy in an infinite lattice medium.

(a) Plane-phase front electromagnetic energy of arbitrary polarization propagating at any angle in the $x-y$ plane in an infinite lattice medium may be decomposed into E plane polarized energy propogating one component in the x direction (with a plane-phase front) and one component in the y direction. Similarly, the H plane portion of the energy can be decomposed into energy propagating with plane-phase fronts in the x and y directions.

(b) One of the above components is called a Bloch wave. We shall speak of that Bloch wave which propagates in the positive y direction with an average lattice propagation constant K_T and an amplitude distribution in the x direction across the cell entrance face given by the function

$$\pi_z = \psi e^{i(K_T y - \omega t)}, \tag{2.16}$$

where π_z is related to the field vectors by Eqs. (2.9) and (2.10).

(c) At successive entrance faces of cells the form of ψ. is maintained as a result of the infinite mutual interactions on either side of the cell under consideration. Therefore, if in an infinite medium there exists K_1 and $K_1 + (2\pi/l)$ (2 Bloch waves) at successive cell entrance faces, both will be maintained and propagated in the *infinite* lattice medium even when the cell dimension m is less than the free space λ.

(d) In progressing through the first half of the cell the E and H fields are distorted to fit the mode solutions of the obstacle. The larger the obstacle the greater the distortion and the greater K_T differs from k_T. In progressing through the second half of the cell the E and H fields are restored.

(e) In general, the lattice medium is dispersive, i.e., the index of refraction, $N = K_T/k_T$, is a function of the frequency,

$$f = \frac{3\times 10^8}{\lambda}; \quad k_T = \frac{2\pi}{\lambda}.$$

Only in special cases, when the obstacle is very small (ψ substantially constant throughout the cell) compared to cell size and λ is very much larger than cell size, will N be almost constant with respect to f. This is the case where the lattice medium may be replaced by an equivalent homogeneous medium of relative dielectric constant $= N^2/\mu$rel. (N is very close to 1 in this case).

(f) The difference in the geometrical shape of the mode solutions for the obstacle from that of the cell boundary introduces the pass and stop bands illustrated in Fig. 4. More basically, it may be said that the spacing, arrangement, and shape of the obstacles making up the lattice medium cause it to have a wave band pass filter type of propagation characteristic.[10] It is interesting to note that when λ approaches infinity, the "fine-grain" structure of the lattice becomes unimportant and it appears to be a homogeneous dielectric[14] in the E plane case and a homogeneous, perfectly conducting parallel plate wave guide in the H plane case.[31] Also as λ approaches infinity, for finite obstacle size, the interactions become less and less important and "far-field" conditions are reached at each cell boundary independently. The "fine-grain" structure predominates, and the average lattice propagation constant K_T is very nearly equal to k_T and has very little meaning since the energy is propagated at the free-space velocity and the phase velocity in the far field is also that of the speed of light.

(g) The forbidden or stop bands of k are associated with point values of $K = (2n+1)\pi/l$, where l = cell length in the direction of propagation. This means that there are a series of discrete values of the average lattice propagation constant K associated with Bloch waves at which stable transmission in the infinite lattice medium cannot occur. In the finite lattice structure these values correspond to the Bragg reflections in crystals.

3. *Impedance Concepts—Lattice Structures*

In the preceding section we have discussed the propagation constant of the infinite lattice medium. We have found that the Wigner-Seitz cellular method of approach is a very powerful and basic one when dealing with infinite lattices. When dealing with lattice structures which have a gross and face perpendicular to the Y axis starting somewhere along the y axis other than at minus infinity (but are effectively infinite in extent in the $\pm x$ and $\pm z$ directions), we find that in addition to knowing the propagation constant of the medium one must also know something about the gross and effect of

the medium. This is analogous to the one-dimensional transmission line. When it is doubly infinite, one need only know its propagation constant. When it has a pair of entrance terminals at other than $y=-\infty$ one must know, in addition, its characteristic impedance (i.e., the input impedance of a semi-infinite length of line) in order to completely specify the energy transmission across the input terminal and thence down the semi-infinite line. Even if the line is not semi-infinite, if its effective characteristic impedance is connected across the output, the energy transmission at the input terminals is specified.

Thus we will attempt to specify a characteristic impedance Z_c for our over-all finite lattice structure which coupled with the already calculated propagation constant K_T will permit the specification of the matrix impedance coefficients of an equivalent 4 pole which may be used as a component of an over-all microwave transmission system. (The coefficients will be functions of the various transmission and impedance modes defined for our "wave-filter" lattice structure.)

Referring to Fig. 5, we have for an over-all length of lattice$=nl$ ($n=$number of elements in the y direction and $l=$length of unit cell in y direction),

$$\begin{aligned} Z_{aa}&=Z_{bb}=Z_c\coth(nilK_T)=Z_c\cot(nlK_T),\\ Z_{ab}&=Z_{ba}=Z_c\,\mathrm{csch}\,(nilK_T)=Z_ci\csc(nlK_T). \end{aligned} \tag{2.17}$$

Referring back to our discussion of the effect of mutual interaction, we can qualitatively discuss what happens to the E and H fields as the energy enters at the finite input face of the lattice structure (infinite in the x and z directions).

That portion of the energy which is transmitted at the first row of cells is acted upon by infinite mutual interactions from the $+y$ direction and none from the $-y$ direction. This causes a change in the amplitude distribution (over the x direction) of the wave on the cell exit face as compared to the cell entrance face, but the direction of propagation and plane phase front is maintained at each cell face. As the wave proceeds farther into the medium, the residual amplitude distortion from entrance to exit face becomes less and less, until finally it has gone deeply enough into the lattice structure so that for all practical purposes the interactions from the $+y$ and $-y$ directions are equal; from then on that particular amplitude distribution is maintained until the exit transition region is reached.

The exact calculation of the E and H fields in the transition regions is very complicated and would require the application of boundary conditions differing from Eq. (2.4). For our purposes it will be sufficient to assume an over-all amplitude distortion caused in the plane wave incident upon the lattice, and use the new amplitude distribution over the entrance face of the first cell past the transition region. An integral function (calculated at this point on the basis of Poyntings vector and using the expressions for π_s already derived) will be our definition of the characteristic impedance. It might be argued that our π_s was calculated on the basis of interaction from $y=-\infty$ to $y=+\infty$ which is not the condition existing at the interior entrance face of the finite lattice where we have decided to calculate our characteristic impedance and propagation constant. This is true, but the discrepancy is not very serious. It is the same one which exists when one neglects high frequency leakage unbalance in breaking off a doubly infinite line of 4 poles and assigning a characteristic impedance value to the remainder which was actually calculated on the doubly infinite basis.

The field impedance concept developed by Schelkunoff is not only a good engineering tool but also has deep significance in terms of electromagnetic phenomena. It takes the $\bar{E}$ and $\bar{H}$ fields which are, in general, functions of position in three-dimensional space and so integrates and interprets them as to permit their energy transmission properties to be represented by network impedances lumped in at least the cross-sectional dimensions. In our case, because we are able to determine the propagation constant of the lattice structure by other means, we integrate in the third dimension also and come out with one equivalent 4 pole (i.e., one in form, but many—corresponding to the various modes—in detail).

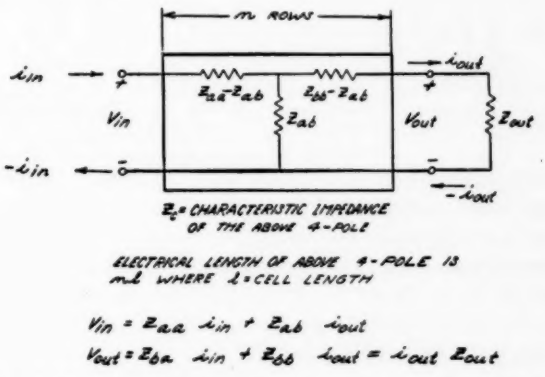


FIG. 5. Network representation of n rows in the interior of a lattice medium.

An example of the basic nature of the Schelkunoff field impedance concept is furnished by the case of the perfectly conducting two-wire transmission line with the wires spaced a finite distance apart. If one takes the attitude that the energy is transmitted in the wires it might be a little difficult to understand why the characteristic impedance of such an infinite line is finite. However, if one takes the attitude that the energy is transmitted in the dielectric space between and surrounding the two wires and that the perfect conductors merely act as coalescers or guides for the energy, then it becomes obvious that the energy propagation is finite and somewhat slower than in vacuum because of the constricting effects of the two conductors. Thus a calculation of field intrinsic impedance based on the Poynting vector integration will very logically and

neatly result in a characteristic impedance value keyed to the energy transmission capabilities of the configuration.

Let us define the Poynting vector for the field vectors defined in Eqs. (2.8) through (2.11) as the power crossing an entrance face in the infinite lattice medium:

$$S_y=\int_{x=-m/2}^{m/2}(\bar{E}\times\dot{H})\cdot n_y' dx=I_q\dot{I}_q Z_c, \quad (2.18)$$

where $\dot{I}_q$ = conjugate I_q, $\dot{H}$ = conjugate $\bar{H}$, and I_q denotes an integrated current caused by the $\bar{H}$ distribution over the entrance face. Z_c = the lattice characteristic impedance.

For the E plane case we have

$$S_y^*=\int_{x=-m/2}^{x=m/2}(-E_x^*\dot{H}_z^*)dx=I_{qp}\dot{I}_{qp}^* pZ_c^*, \quad (2.19)$$

where p denotes the impedance mode under consideration:

$$I_q^*=\int\int H_z^*\cdot d\bar{s}=H_{zq}^*=k^2\pi_{zq}^* =k^2\int_{x=-m/2}^{m/2}\pi_z^* dx, \quad (2.20)$$

where the path in Eq. (2.20) is in accordance with the right-hand rule for I_q^* in the positive y direction and over a 1 meter unit length in the z direction (see Fig. 6). Remembering that H_z^* is constant with respect to Z and that $H_z^*\cdot dx=0$, we see that I_q^* is numerically equal to H_{zq}^* where the "q" denotes evaluation of current at $x=0$, $y=-l/2$ and $H_{zq}^*=2(H_z^*)$:

$$S_y^*=H_{zq}^*\dot{H}_{zq}^* Z_c^*, \quad (2.21)$$

$$pZ_c^*=\frac{S_y^*}{H_{zq}^*\dot{H}_{zq}^*}=\frac{\int_{x=-m/2}^{m/2}(-E_x^*\dot{H}_z^*)dx}{H_{zq}^*\dot{H}_{zq}^*}, \quad (2.22)$$

where

$$E_x^*=\frac{iZ_0^*}{k}\frac{\partial H_z^*}{\partial y}=ikZ_0^*\frac{\partial\pi_z^*}{\partial y},$$

$$pZ_c^*=\frac{\int_{x=-m/2}^{m/2}\left[-ik\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\partial}{\partial y}\pi_z^* k^2\dot{\pi}_z^*\right]dx}{k^4\pi_{zq}^*\dot{\pi}_{zq}^*}. \quad (2.23)$$

Normalizing Eq. (2.23) by the characteristic impedance of free space per cell face area of m sq meters we get

$$\frac{pZ_c^*}{(mZ_0^*)}=\frac{1}{ik(m)}\left(\frac{1}{\pi_{zq}^*\dot{\pi}_{zq}^*}\right)\times\int_{x=m/2}^{m/2}\left(\frac{\partial}{\partial y}\pi_z^*\right)\dot{\pi}_z^* dx. \quad (2.24)$$

But at the cell entrance face where Z_c^* is being calculated we have $y=-l/2$ and from the cell entrance and exit face boundary conditions ψ_T^* is not a function of y (if we restrict our sampling points to cell end faces), therefore,

$$\pi_z^*=(e^{iK_T^* y})\psi_T^*, \quad (2.25)$$

$$\frac{\partial}{\partial y}\pi_z^*=iK_T^*\pi_z^*. \quad (2.26)$$

In the pass bands T, K_T is real (whereas in the stop bands it is imaginary):

$$\frac{pZ_c^*}{(mZ_0^*)}=\frac{K_T^*}{mk_T}\left(\frac{1}{\pi_{zq}^*\dot{\pi}_{zq}^*}\right)\times\int_{x=-m/2}^{m/2}\left[\left(\pi_z^*\frac{e^{iK_T^* y}}{K_T^*}\frac{\partial\dot{\psi}^*}{\partial y}\right)+(\pi_z^*\dot{\pi}_z^*)\right]dx, \quad (2.27)$$

Eq. (2.27) is less than or equal to K_T^*/k_T.

We see that for the special case of π_z^* constant with respect to x, y, and z we get

$$\frac{pZ_c^*}{(mZ_0^*)}=\frac{mK_T^*}{mk_T^*}=N_T^*=\text{index of refraction}. \quad (2.28)$$

Equation (2.28) is the index of refraction of a homogeneous dielectric of absolute dielectric constant $=(K_T^*)^2/\mu_0$.

For this special case the general voltage reflection coefficient, which is

$$\text{Reflection coeff}=\frac{[Z_c^*/(mZ_0^*)]-1}{[Z_c^*/(mZ_0^*)]+1}, \quad (2.29)$$

becomes

$$(N^*-1)/(N^*+1) =\text{reflection coeff for homogeneous dielectric}. \quad (2.30)$$

It should be noted that in the preceding equations Z_c^* varies with the transmission mode being considered so that properly we should include "T" as a subscript and

$$\frac{pZ_{TC}^*}{(mZ_0^*)}=\frac{N_T^*}{m}\left(\frac{1}{\pi_{zq}^*\dot{\pi}_{zq}^*}\right)\times\int_{x=-m/2}^{m/2}\left[\left(\pi_z^*\frac{e^{iK_T y}}{K_T^*}\frac{\partial\dot{\psi}^*}{\partial y}\right)+(\pi_z^*\dot{\pi}_z^*)\right]dx. \quad (2.31)$$

Since K_T^* is already an averaged value in so far as x is concerned, we may rewrite Eq. (2.31) using Eq. (2.2):

$$\frac{pZ_{TC}^*}{(mZ_0^*)}=\frac{N_T^*}{m}\left(\frac{1}{\psi^*(q,k_T)\dot{\psi}^*(q,k_T)}\right)\times\int_{x=-m/2}^{x=m/2}\left[\left(\psi^*\frac{1}{K_T^*}\frac{\partial\dot{\psi}^*}{\partial y}\right)+(\psi^*(x,k_T)\dot{\psi}^*(x,k_T)\right]dx. \quad (2.32)$$

If one were interested only in the total reflection characteristics at the input end of a finite length of lattice medium for a given impressed k, and if one were not too concerned about loss of available output energy as a result of deformations of phase and amplitude fronts at the output end caused by the energy being transmitted by more than one Bloch wave transmission mode, it would not be necessary to use Eq. (2.32). Instead, one would use Eq. (2.21). In order to take care the effects of the transition region, one would assume that it caused a residual variation of the incident H_z^* along the entrance face past that region, consistent with the lattice symmetry requirement that this distribution be an even function of x (about the $x=0$ point) and with the requirement that no energy be propagated across the cell side faces. Since

$$H_{zq}^*\dot{H}_{zq}^* = |H_z^*|^2_{\text{at } x=0}, \tag{2.33}$$

and for measurements at cell end faces,

$$\frac{\partial}{\partial y}H_z^* = iKH_z^*. \tag{2.33a}$$

Therefore, we may express the characteristic impedance in terms of a mean square value of the distribution along the x axis. If the distribution along the x axis is written in terms of a Fourier series,

$$H_z^*(x) = \sum_{p=0}^{\infty}\left(A_p\cos\frac{2\pi p}{m}x + B_p\sin\frac{2\pi p}{m}x\right),$$
$$H_{zq}^*\dot{H}_{zq}^* = \int_{x=-m/2}^{m/2} H_z dx \int_{x=-m/2}^{m/2} \dot{H}_z dx. \tag{2.34}$$

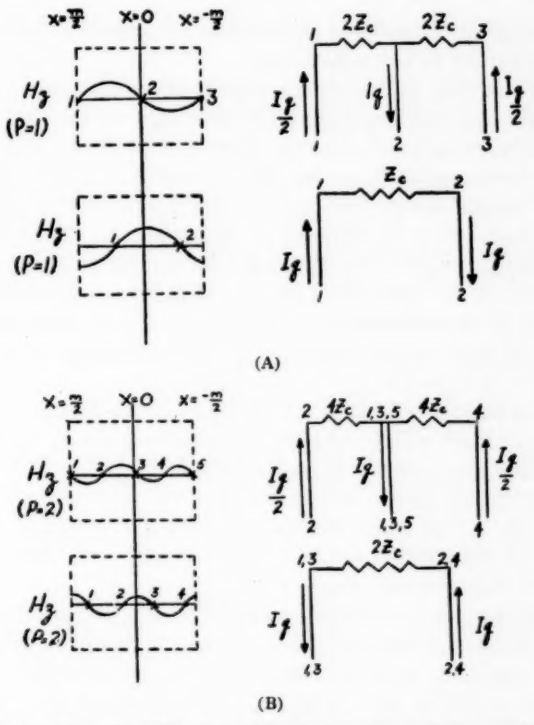
In terms of power transmission one may define a space form factor F_S^* which, when multiplied with the complex conjugate product of the H_z^* field at $x=0$ and $[(K/k)mZ_0^*]$, will give the proper total value of the Poynting vector S_y^*.

$$\therefore\quad F_S^* = \frac{1}{m}\frac{\int_{x=-m/2}^{m/2}(H_z^*\dot{H}_z^*)dx}{H_{zq}^*\dot{H}_{zq}^*}$$
$$= \frac{1}{K_T^*}\frac{\int_{x=m/2}^{m/2}\left(\frac{\partial \pi_z^*}{\partial y}\right)\dot{\pi}_z^* dx}{\pi_{zq}^*\dot{\pi}_{zq}^*}. \tag{2.35}$$

$$\therefore\quad S_y^* = [F_S^* H_{zq}^* H_{zq}^*]\frac{K^*}{k^*}(mZ_0^*). \tag{2.36}$$

Substituting Eq. (2.36) into Eqs. (2.22) we get

$$\frac{Z_{TC}^*}{mZ_0^*} = F_S^*\frac{K_T^*}{k_T} = F_S^* N_T^*. \tag{2.37}$$



FIGS. 6(A) and 6(B). Sinusoidal distributions of H_z amplitude modes across the cell entrance face and equivalent network representation of each. In (A) points 1 and 3 may be tied together; in (B) points 2 and 4 may be tied together.

(Note: The form factor does not take into account rms values in the time domain, so we consider that the Poynting vector S_y^* and the field vectors are given in terms of rms time values). In Eq. (2.37) the consideration of the transmission modes enters only in the choice of N_T^* corresponding to the particular k_T of the impressed frequency as given by Fig. 4. Approximations of F_S^* may be found by numerical integration of the formulas for π_z^* and $\partial\pi_z^*/\partial y$ which are derived in Sec. III.

On the other hand, if one is interested in the destination of the energy and in the detailed phase front information at the output end, an impedance mode analysis of the energy transmission must be made. An analysis is required which will show how many different Bloch waves modes will propagate the energy through the lattice when a plane wave of frequency ω_0, $(k_0=\omega_0/3\times10^8)$ impinges upon the entrance face of the lattice.

If the finite face of the lattice (at the point past the entering transition region) were replaced by a homogeneous dielectric, periodic in the x direction with a period of m, the condition for a propagating mode to exist is given by

$$\left(\frac{2\pi n}{m}\right)^2 \leq k_0^2 \tag{2.38}$$

(where n is any integer from zero to infinity and is the number of the propagating mode resulting from an

impinging plane wave constant in amplitude across the wave front in the x direction).

Thus, Eq. (2.38) shows that if $m<\lambda$ only the $n=0$ mode will propagate past the boundary of the periodically varying homogeneous medium. For $\lambda<m<2\lambda$ both the $n=0$ and $n=1$ modes will propagate energy through the homogeneous dielectric. Using this type of analysis we may reason that in Eq. (2.31) or Eq. (2.32) an incident plane wave with free-space wave number k_1 located in the "Tth" pass band as defined in Fig. 4 will propagate all its transmitted energy in one Bloch wave mode K_T defined as the value of K corresponding to k_1 in Fig. 4 if $m/\lambda_0<1$.

On the other hand if $1<(m/\lambda_0)<2$, the energy to be transmitted will be shared in two Bloch wave modes, namely, the one corresponding to k_1 in Fig. 4 and the next higher mode which is equal to $K_{T0}+(2\pi/l)$. If $2<(m/\lambda_0)<3$, the energy transmitted through the lattice is shared in three Bloch wave modes—the preceding two and $K_{T0}+4(\pi/l)$.

In the cases where the energy is transmitted through the lattice by means of more than one Bloch wave and one wishes to keep account of the separate modes, one must calculate the characteristic impedances $Z_{TC}{}^*$ appropriate for each of the transmitting Bloch wave modes and set up matching boundary conditions between the distorted incident plane wave existing at the first lattice cell face past the transition region and the sum of the n transmitting Bloch waves determined as outlined above.

When $\lambda_0>m$, only one Bloch wave mode (corresponding to k_1 in Fig. 4) will be transmitted and only one 4 pole with $Z_{TC}{}^*$ from Eqs. (2.37) or (2.32) need be used in the microwave transmission circuit representation. If more than one Bloch wave mode propagates the energy, but the energy in the higher Bloch wave modes transmitting is not considered to be available for use at the output, one may approximate the nonavailability of energy by shunting the input to the 4 pole calculated from the lowest propagating Bloch wave mode with conductances

$$G_1=\frac{1}{Z_{TC1}{}^*},\quad G_2=\frac{1}{Z_{TC2}{}^*},\text{ etc.}$$

We shall now examine the criterion of Eq. (2.38) a little more closely. As was mentioned previously, we are interested in the lattice as an electromagnetic energy transmitter. Energy which is obliquely incident to the y axis may be handled by means of the *PMNQ* and *USTV* functions developed in Sec. III, provided the effect of the transition region is simulated by a function analogous to the $F_S{}^*$ function of Eq. (2.35) which serves to reduce the amplitude of the incident wave by the amount resulting from Bragg[7] reflections. One must also suitably change the direction of the normal to the plane constant phase front penetrating the lattice medium and suitably alter the amplitude distribution along the phase front caused as the energy proceeds through the transition region. Once the plane constant phase front has reached the region of balanced interactions, the wave may be divided into two waves—one proceeding along the y axis and the other along the x axis. The former handled as in the preceding paragraphs coupled with the *PMNQ* functions of the next section and the latter handled in an entirely analogous manner but using the x axis as the direction of energy transmission coupled with the *USTV* functions. However, care must be exercised at the finite output end (before the output transition region) to properly combine the energy flows into one wave of proper total direction and amplitude distribution.

Now let us establish Eq. (2.38) on the basis of energy normally incident in the y direction only. Since the lattice medium is considered to be finite in the y direction but infinite in the $x-z$ plane, the direction of propagation of the normally incident energy is not changed when it arrives at the first cell entrance face layer past the transition region. From this point on the energy will proceed with a propagation constant $K_T{}^*$. At this point the distorted amplitude distribution in the x direction must be periodic as illustrated in Fig. 7.

$$\pi_z{}^*=\sum_{p=0}^{\infty}\left[A_p\cos\left(\frac{2\pi p}{m}x\right)+B_P\sin\left(\frac{2\pi p}{m}x\right)\right]\psi_e{}^* iK_T y. \quad (2.39)$$

Also at this point $\pi_z{}^*$ must still satisfy the scalar wave equation in free space.

Therefore, $\nabla^2\pi_z{}^*+k^2\pi_z{}^*=0$ or

$$\frac{\partial 2\pi_z{}^*}{\partial x^2}+\frac{\partial 2\pi_z{}^*}{\partial y^2}=-k^2\pi_z{}^*. \quad (2.40)$$

Where $k=2\pi/\lambda$, and where $\lambda=$free-space wavelength. Substituting Eq. (2.39) into Eq. (2.40) we get

$$-\left(\frac{2\pi p}{m}\right)^2\pi_z{}^*-K_T{}^2\pi_z{}^*=-k^2\pi_z{}^*, \quad (2.41)$$

$$K_T{}^2=k^2-\left(\frac{2\pi p}{m}\right)^2. \quad (2.42)$$

Since K_T must be real, we have

$$(2\pi p/m)^2\leq k^2. \quad (2.43)$$

If $\lambda>m$ then Eq. (2.43) is satisfied only if $p=0$. Therefore, for $\lambda>m$ only the Bloch wave mode appropriate to the value of $k=\omega/V_{air}=2\pi f/V_{air}$ (where f is the impressed frequency) will be propagated through that part of the lattice where the equivalent of infinite interactions occur. If $2\lambda>m>\lambda$, then both $p=0$ and

$p=1$ modes may propagate energy. If $3\lambda>m>2\lambda$, then three modes, $p=0$, $p=1$, and $p=2$, may propagate energy.

Thus Eqs. (2.31) or (2.37) may be written with the characteristic impedance showing subscripts for both the transmission modes and impedance modes. Therefore we would write $Z^*_{(T_p)C}$ to denote the characteristic impedance for the pth impedance mode of the Tth transmission mode. The approximate Bloch wave propagation constant for the pth propagating impedance mode would be given by

$$K_{T(p)}=[K_{T0}+(2\pi p/l)]. \tag{2.44}$$

Now let us consider the transmission of energy through the lattice and the storage of energy within the lattice. We will use the concept of the Poynting vector. The complex Poynting vector is given by

$$\int_{y=-l/2}^{l/2}\int_{x=-m/2}^{m/2}(E^*_{\text{Total}}x\dot{H}_{\text{Total}})dxdy \cdot\bar{E}^*_{\text{Total}}x\dot{H}^*_{\text{Total}}=(-E_x{}^*\dot{H}_z{}^*)_{y\text{ direction}}+(E_y{}^*\dot{H}_z{}^*)_{x\text{ direction}}. \tag{2.45}$$

Therefore, the reactive power at cell entrance face is

$$S_{xp}{}^*=\int_{x=-m/2}^{m/2}E_y{}^*\dot{H}_z{}^*dx=\int_{x=-m/2}^{m/2}-ik^3Z_0{}^*\frac{2\pi p}{m}\times\sum_{p=1}^{\infty}\left[B_p\cos\frac{2\pi p}{m}x-A_p\sin\frac{2\pi p}{m}x+\left(A_p\cos\frac{2\pi p}{m}x+B_p\sin\frac{2\pi p}{m}x\right)\right]dx. \tag{2.46}$$

It is seen that for the principal mode (i.e., when $p=0$) there is no reactive power distribution because $E_y{}^*$ is identically zero over the entire cell entrance face. However, at other than the entrance and exit faces this is not so:

$$\begin{aligned}S_{xp}{}^*&=-ik^3Z_0{}^*\frac{2\pi p}{m}\int_{x=-m/2}^{m/2}A_pB_p\left(\cos^2\frac{2\pi p}{m}x-\sin^2\frac{2\pi p}{m}x\right)dx\\&\quad+\int_{x=-m/2}^{m/2}(B_p{}^2-A_p{}^2)\cos\frac{2\pi p}{m}x\sin\frac{2\pi p}{m}xdx\\&=-ik^3Z_0{}^*\frac{2\pi p}{m}A_pB_p\int_{x=-m/2}^{m/2}\cos\frac{4\pi p}{m}xdx\\&=-ik^3Z_0{}^*A_pB_p\left[\sin\frac{4\pi p}{m}x\right]_{-m/2}^{m/2}=0.\end{aligned} \tag{2.47}$$

Therefore, the total reactive power should be found by a double integration along both x and y (rather than just at $y=l/2$) in order to find the equivalent of Cohn's shunt impedance. (Of course, the maximum reactive, or storage, field effect occurs at the $y=0$, $-m/2<x<m/2$ face.)

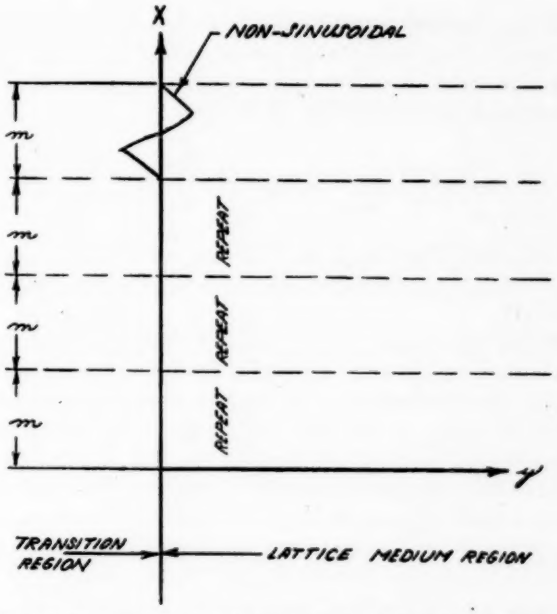


FIG. 7. Illustration of periodicity in the x direction with period of length "m." (Distribution across each cell entrance face in a given row is not, in general, sinusoidal.)

Since we know ψ^* at the face $y=0$, $-m/2<x<m/2$, we can get an approximate value for reactlve power:

$$S^*_{x\text{ reactive}}=\int_{x=-m/2}^{m/2}(E_y{}^*\dot{H}_z{}^*)_{y=0}dx=ik^3Z_0{}^*\left|\int_{x=-m/2}^{m/2}\left(\frac{\partial\pi_z{}^*}{\partial x}\dot{\pi}_z{}^*\right)_{y=0}dx\right|. \tag{2.48}$$

(Remember that $E_y{}^*=ikZ_0{}^*(\partial\pi_z{}^*/\partial x)$ and $H_z{}^*=k^2\pi_z{}^*$.)

It should be noted that this integration goes through the obstacle but since $\bar{E}^*$ and $\bar{H}^*$ are zero in the obstacle the integration is from $x=-m/2$ to $x=-\epsilon_1$ and from $x=\epsilon_1$, to $x=m/2$, where ϵ_1 is the coordinate of the obstacle in the $x-z$ plane at $y=0$. The same impedance mode criteria Eq. (2.44) as for the real Poynting vector hold for the reactive Poynting vector Eq. (2.48).

For the principal wave ($p=0$) we get Cohn's[20] value for shunt impedance if we take $S^*_{x\text{ reactive}}=I_q{}^*\dot{I}_q{}^*Z_S{}^*$:

$$Z_S{}^*=\frac{S^*_{x\text{ reactive}}}{I_q{}^*\dot{I}_q{}^*}=\frac{S^*_{x\text{ reactive}}}{\left[\int\int\bar{H}^*\cdot d\bar{s}\right]^2}\simeq\frac{S^*_{x\text{ reactive}}}{k^4\int_{x=-m/2}^{m/2}(|\pi_z{}^*|)^2dx}. \tag{2.49}$$

It is seen that $Z_S{}^*$ is a pure reactance.

If the obstacle is small enough so that π_z^* is virtually constant along the planes $y=0$; $-m/2<x<m/2$ and the side faces, we may make a further approximation resulting in a value of Eq. (2.48) of

$$(S^*_{x\,\text{reactive}})_{\text{approx}} = -ik^3 Z_0^* m \left| \left(\frac{\partial \pi_z^*}{\partial x} \cdot \dot{\pi}_z^* \right) \right|. \quad (2.50)$$

In Eq. (2.50) $\dot{\pi}_z^*$ and $\partial \pi_z^*/\partial x$ are evaluated at $y=0$, $x=-m/2$ using the formulas which will be developed in Sec. III.

While this paper is not primarily concerned with the transition region, it is believed that the foregoing point of view may be used to advantage in performing a more detailed study of the transition region. As a first approximation we may assume that the k *vs* K characteristic of the transition zone is the same as that for the infinite lattice medium. The principal effects, then, as the normally incident plane wave passes successive cell entrance faces is a decrease in amplitude of the E and H vectors and a residual distortion of the E and H distribution at the cell exit face as compared to the distribution at the cell entrance face.

Within any given transmission band, the discontinuity introduced by the transition region may be replaced by an equivalent storage field. This storage field, in turn, may be represented by one, or a series of, lumped circuit parameters in tandem or by a distributed parameter circuit. Past the transition region the semi-infinite lattice medium may be represented by the lumped parameter four pole of Eq. (2.17).

Since the worst portion of the discontinuity attributed to the transition region occurs at the first layer or grating of a lattice structure, we may approximate the transition region by a suitable increase in the lumped impedance parameter of the single grating as calculated by Marcuvitz and others. For example, if the impedance for a single grating is a single 4-pole, and we assume that the transition region of the lattice structure consists of six layers or gratings, the discontinuity effect of the transition region is less than six times the single discontinuity effect. As a first approximation we may assume that the portion of the transition past the first layer acts as a tapered matching network between the single 4-pole representing the initial layer or grating and the 4-pole representing the semi-infinite medium of Eq. (2.17).

As a second approximation one might want to take into account the fact that the transition region itself acts as a selective frequency filter whose responses are not identical with that of the infinite lattice medium. One way of taking this into account is to modify the boundary conditions of Eq. (2.4) for the calculation of the transcendental equation governing the k *vs* K_T frequency characteristic. Thus we get

$$\begin{aligned} \psi^*_{\text{entrance}} &= \psi^*_{\text{exit}} e^{i\theta} e^{iK_T l}, \\ \hat{n} \cdot (\nabla \psi^*_{\text{entrance}}) &= e^{i\theta} e^{iK_T l} [\hat{n} \cdot (\nabla \psi^*_{\text{exit}})]. \end{aligned} \quad (2.51)$$

In Eq. (2.51) θ is complex, in general, and is a measure of the lack of full restoration of the wave at the exit face as compared with the entrance face. θ is a function of distance from the initial layer of the lattice structure. The magnitude of θ may be considered to be an exponential decay starting at a certain value at the first layer and reaching a negligible value at the last layer of the transition region. In the simple case where the transition region is just one or two layers thick we might represent θ by

$$\theta = A\,(\exp) - (y+l/2)/l. \quad (2.52)$$

Obviously the resulting relationship between k and K_T is much more complicated than the transcendental Eq. (2.15).

Similar analyses may be carried out for the H plane case. Here we have

$$\begin{aligned} E_z &= k^2 \pi_z, \\ H_x &= -ik \frac{1}{Z_0} \frac{\partial \pi_z}{\partial y}, \\ H_y &= ik \frac{1}{Z_0} \frac{\partial \pi_z}{\partial x}. \end{aligned} \quad (2.53)$$

Thus we obtain the relationships

$$\int_{x=-m/2}^{m/2} (E_z \dot{H}_x) dx = \frac{E_q \dot{E}_q}{Z_{TC}}, \quad (2.54)$$

$$E_q = \left[\int\int \bar{E} \cdot d\bar{s} \right] = E_{zq} = \int_{x=-m/2}^{m/2} k^2 \pi_z dx \simeq m k^2 \pi_z. \quad (2.55)$$

This leads to the following equations:

$$\begin{aligned} \frac{Z_{TC}}{Z_0} &= \frac{k^4 \int_{x=-m/2}^{m/2} \pi_z dx \int_{x=-m/2}^{m/2} \dot{\pi}_z dx}{-ik^3 \int_{x=-m/2}^{m/2} \pi_z \frac{\partial \dot{\pi}_z}{\partial y} dx} \\ &= ik \frac{\int \pi_z dx \int \dot{\pi}_z dx}{\int \pi_z \frac{\partial \dot{\pi}_z}{\partial y} dx}, \end{aligned} \quad (2.56)$$

$$\begin{aligned} \frac{Z_{TC}}{Z_0} &= \frac{k}{K_T} \frac{\int \pi_z dx \int \dot{\pi}_z dx}{\int [\pi_z \dot{\pi}_z + \pi_z (e^{ikv}/K_T)(\partial \psi / \partial y)] dx} \\ &= \frac{1}{F_s N_T}, \end{aligned} \quad (2.57)$$

$$S_z=\int_{y=-l/2}^{l/2}\int_{x=-m/2}^{m/2}-(E_z\dot{H}_y)dxdy=\frac{E_q\dot{E}_q}{Z_S}. \tag{2.58}$$

Eliminating the integration along the y axis by using the maximum values at $y=0$, we get

$$\frac{Z_S}{Z_0}=\frac{k^4\int_{x=-m/2}^{m/2}\pi_z dx\int_{x=-m/2}^{m/2}\dot{\pi}_z dx}{-ik^3\int_{x=-m/2}^{m/2}\pi_z(\partial\dot{\pi}_z/\partial x)dx}$$

$$=ik\frac{\int\pi_z dx\int\dot{\pi}_z dx}{\int\pi_z(\partial\dot{\pi}_z/\partial x)dx}. \tag{2.59}$$

This section has considered fundamental propagation and impedance concepts pertinent to all two-dimensional lattice media and structures. In the next section we will specialize our discussion to lattice media composed of elliptic and circular cylinders.

SECTION III

Lattice Media

Mathematical Treatment for Elliptic and Circular Cylinders

1. Introduction.—In this section we will consider a lattice medium of elliptic cylinders for which a unit cell is as shown in Fig. 8. By treating the general case using Mathieu functions, we automatically cover many other geometries which are limiting cases of the elliptic cylinder. Thus, by making the major and minor axes equal we obtain the conventional circular cylinder lattice medium; by shrinking the minor axis to zero, we secure a strip obstacle type of lattice medium; by increasing the major axis of the obstacle to equal, or almost equal, the cell length m we can get the metal plate lens or a coupled plate wave-guide lens or a ridged metal plate lens.

In the succeeding paragraphs we shall apply the Wigner-Seitz cellular method to the solution of wave propagation in a lattice medium of elliptic cylinders.

2. Propagation function—Y direction.—In the following development, we shall use the notation and types of Mathieu functions and Hertz vectors outlined by Stratton.[61]

Referring to Fig. 8,

$C_0=\frac{1}{2}$ distance between foci,
$a_0=$ semi-major axis$=C_0\epsilon_0=C_0\cosh u_0$,
$b_0=$ semi-minor axis$=C_0(\epsilon_0^2)^{\frac{1}{2}}-1=C_0\sinh u_0$,
$1/\epsilon_0=C_0/a_0=$ eccentricity$=e$,
$\epsilon=\cosh u$; $\eta=\cos\nu$,
$x=C_0\cosh u\cos\nu$; $y=c_0\sinh u\sin\nu$,

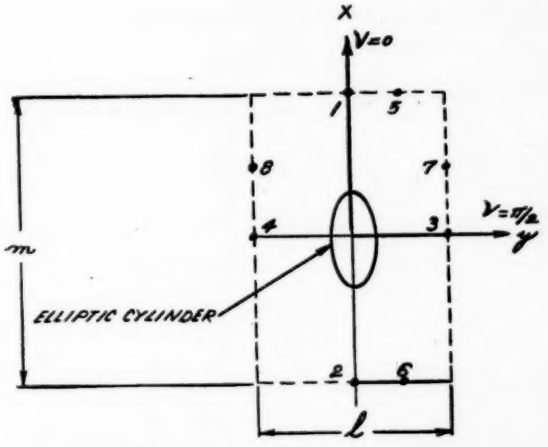


FIG. 8. Diagram of unit cell illustrating points on cell boundaries where boundary conditions are matched.

when $\epsilon\to\infty$, $C_0\to 0$ and $c_0\cosh u\to r$ and $\nu\to\theta$ which are the polar coordinates for the circular cylinder case. When $\epsilon\to 1$, $2c_0\to w$ which describes a thin strip of width w.

$$\frac{x^2}{\epsilon^2}+\frac{y^2}{\epsilon^2-1}=C_0^2,$$
$$\frac{x^2}{\eta^2}-\frac{y^2}{1-\eta^2}=C_0^2. \tag{3.1}$$

We shall now derive the statement of the boundary conditions used in the Slater[50] approximation of the Wigner-Seitz cellular method, considering that more than the usual minimum 2 pairs of points on the unit cell boundary are to be used to match boundary conditions. The physical facts to be applied are that tengential components of $\bar{E}$ and $\bar{H}$ are continuous across the cell boundaries and repetitive at corresponding points on cell boundaries and that the tangential component of $\bar{E}$ and the normal component of $\bar{H}$ vanish on the surface of the perfectly conducting cylinders.

The solutions will be secured in terms of a scalar wave function and Hertz vectors π (where the E plane case is denoted by an asterisk and has $E_z^*=H_x^*=H_y^*=0$ and $H_z^*=k^2\pi_z^*$; the H plane case has $H_z=E_x=E_y=0$ and $E_z=k^2\pi_z$):

$$\pi_z=\exp[i(Ky-\omega t)]\psi=\exp[i(Ky-\omega t)]$$
$$\times\sum_m[A_mS_{em}(R_{em}'+C_mR_{em}^2)$$
$$+B_mS_{0m}(R_{0m}'+D_mR_{0m}^2)]. \tag{3.2}$$

Let

$$Z_{em}=R_{em}'+C_mR_{em}^2$$
and
$$Z_{0m}=R_{0m}'+D_mR_{0m}^2. \tag{3.3}$$

C_m and D_m are determined by the boundary conditions

on the surface of the cylinder:

$$C_m=-\frac{R_{em}'(c_0k,\epsilon_0)}{R_{em}^2(c_0k,\epsilon_0)},$$

$$D_m=-\frac{R_{0m}'(c_0k,\epsilon_0)}{R_{0m}^2(c_0k,\epsilon_0)},$$

$$C_m^*=-\frac{\partial/\partial u[R_{em}'(c_0k,\epsilon_0)]}{\partial/\partial u[R_{em}^2(c_0k,\epsilon_0)]}, \tag{3.4}$$

$$D_m^*=-\frac{\partial/\partial u[R_{0m}'(c_0k,\epsilon_0)]}{\partial/\partial u[R_{0m}^2(c_0k,\epsilon_0)]}.$$

Therefore,

$$\psi=\sum_{m=0}^{\infty}A_mS_{em}Z_{em}+B_mS_{0m}Z_{0m}. \tag{3.5}$$

In Eq. (3.5) the S's and Z's are functions of parameters shown in Stratton.[61]

In order to evaluate the A_m's and B_m's, we must match boundary conditions on the unit cell boundaries. The numbers of pairs of points at which the evaluations are made will govern the number arbitrary constants A and B that can be used to represent the wave function. While the actual diffraction phenomena associated with any given elliptic cylinder can best be described in terms of Mathieu functions in an elliptic geometry, the over-all behavior of electromagnetic energy propagated through a lattice medium composed of a rectangularly disposed array of elliptic cylinders is best expressed in Cartesian coordinates. Therefore we must reconcile and correlate the specification of the electromagnetic fields (and their associated Hertz vectors and scalar wave functions) in both Cartesian and elliptic cylinder coordinates.

Let us refer to Fig. 8 and consider the E plane case:

$$H_z^*=h^2\pi_z^*,$$

$$E_\epsilon^*=\frac{i\omega\mu}{C_0\left(\frac{\epsilon^2-\eta^2}{1-\eta^2}\right)^{\frac{1}{2}}}\frac{\partial\pi_z^*}{\partial\eta}\quad\frac{-i\omega\mu}{C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}}\frac{\partial\pi_z^*}{\partial\gamma}, \tag{3.6}$$

$$E_\eta^*=\frac{-i\omega\mu}{C_0\left(\frac{\epsilon^2-\eta^2}{\epsilon^2-1}\right)^{\frac{1}{2}}}\frac{\partial\pi_z^*}{\partial\epsilon}=\frac{-i\omega\mu}{C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}}\frac{\partial\pi_z^*}{\partial u}.$$

But

$$E_y^*=i\omega\mu\frac{\partial\pi_z^*}{\partial x}=(-i\omega\mu)\left(\frac{\partial\pi_z^*}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial\pi_z^*}{\partial\gamma}\frac{\partial\gamma}{\partial x}\right),$$

$$E_x^*=i\omega\mu\frac{\partial\pi_z^*}{\partial y}=(i\omega\mu)\left(\frac{\partial\pi_z^*}{\partial\mu}\frac{\partial u}{\partial y}+\frac{\partial\pi_z^*}{\partial\gamma}\frac{\partial\gamma}{\partial y}\right). \tag{3.7}$$

Using Eqs. (3.1) and (3.6) and remembering that $\pi_z^*=[e^{i(K_y-\omega t)}]\psi^*$ we have

$$E_\epsilon^*=\frac{-i\omega\mu}{C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}}\times\left[e^{iKy}\frac{\partial\psi^*}{\partial\gamma}+(iKC_0\sinh u\cos\gamma)\pi_z^*\right],$$

$$E_\eta^*=\frac{-i\omega\mu}{C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}}\times\left[e^{iKy}\frac{\partial\psi^*}{\partial u}+(iKC_0\cosh u\sin\gamma)\pi_z^*\right], \tag{3.8}$$

$$E_y^*=C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}\left(E_\eta^*\frac{\partial u}{\partial x}+E_\epsilon^*\frac{\partial\gamma}{\partial x}\right),$$

$$E_x^*=-C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}\left(E_\eta^*\frac{\partial u}{\partial y}+E_\epsilon^*\frac{\partial\gamma}{\partial y}\right). \tag{3.9}$$

Using the preceding equations one has established the general boundary conditions for the repetition of tangential E and H fields on corresponding unit cell side faces and end faces, respectively. On the side faces E_y^* and H_z^* are matched at corresponding points 1, 2; 5, 6; etc., while on the end faces E_x^* and H_z^* are matched at corresponding points 3, 4; 7, 8; etc., as shown in Fig. 8. It should be noted that the boundary conditions are given in terms of ψ^*.

We shall now specialize the above to cover the case of 2 pairs of matching points; namely, points 1 and 2 and points 3 and 4 on Fig. 8.

At points 1 and 2,

$$\frac{\partial\gamma}{\partial x}=0,\quad\left(\frac{\partial u}{\partial x}\right)_1=-\left(\frac{\partial u}{\partial x}\right)_2; \tag{3.10}$$

at points 3 and 4,

$$\frac{\partial\gamma}{\partial y}=0,\quad\left(\frac{\partial u}{\partial y}\right)_3=-\left(\frac{\partial u}{\partial y}\right)_4.$$

Therefore, the boundary conditions reduce to

$$\psi_1^*=\psi_2^*,\qquad\frac{\partial\psi_1^*}{\partial u}=-\frac{\partial\psi_2^*}{\partial u},$$

$$\psi_3^*=(e^{-iKl})\psi_4^*,\quad\frac{\partial\psi_3^*}{\partial u}=-e^{iKl}\frac{\partial\psi_4^*}{\partial u}. \tag{3.11}$$

The H plane case ($E_z = k^2\pi_z$) can be treated in an analogous manner and identical boundary conditions will be derived:

$$\psi_1=\psi_2, \quad \frac{\partial\psi_1}{\partial u}=\frac{-\partial\psi_2}{\partial u},$$
$$\psi_3=(e^{-iKl})\psi_4 \quad \frac{\partial\psi_3}{\partial u}=-e^{-iKl}\frac{\partial\psi_4}{\partial u}. \tag{3.12}$$

Therefore, in the succeeding development of the relationship between K and k the form of the formulas will be identical for the asterisked (E plane) case and the unasterisked (H plane) case. However one must remember that the definitions of Z^* and Z given in Eqs. (3.3) and (3.4) are different.

For the case of two pairs of tangential boundary conditions (4 matching equations) ψ may only contain five arbitrary $A-B$ constants, i.e., one more than the number of boundary matching equations. This extra constant takes into account the arbitrary amplitude of the incoming plane electromagnetic wave. In general the number of arbitrary $A-B$ constants will be 1+number of cell boundary matching points.

Therefore, for two pairs of matching points 1–2, 3–4 in Fig. 8, Eq. (3.11) becomes

$$\psi=A_0S_{e0}Z_{e0}+A_1S_{e1}Z_{e1}+A_2S_{e2}Z_{e2} +B_0S_{00}Z_{00}+B_1S_{01}Z_{01}. \tag{3.13}$$

We shall denote differentiation by enclosing the quantity with a parenthesis coupled with a prime superscript. Derivatives of Mathieu functions will be with respect to μ and γ. Thus, for example,

$$(S_{e0})'=\frac{\partial}{\partial\gamma}(S_{e0}); \quad (Z_{e0})'=\frac{\partial}{\partial u}(R_{e0}^{1})+C_0\frac{\partial}{\partial u}(R_{e0}^{2}). \tag{3.14}$$

Applying $\psi_1=\psi_2$, we find

$$A_1=-A_1=0. \tag{3.15}$$

Applying $\partial\psi_1/\partial u=-\partial\psi_2/\partial u$, we find

$$\begin{aligned}&A_0(S_{e0})_1(Z_{e0})_1'+A_2(S_{e2})_1(Z_{e2})_1' \\ &\quad+B_0(S_{00})_1(Z_{00})_1'+B_1(S_{01})_1(Z_{01})_1' \\ &\quad=-A_0(S_{e0})_2(Z_{e0})_2'-A_2(S_{e2})_2(Z_{e2})_2' \\ &\quad-B_0(S_{00})_2(Z_{00})_2'-B_1(S_{01})_2(Z_{01})_2'.\end{aligned} \tag{3.16}$$

Simplifying Eq. (3.16), we get

$$A_0(Z_{e0})_1'+A_2(Z_{e2})_1'=0. \tag{3.17}$$

And thus we arrive at the relationship

$$A_2=-A_0[(Z_{e0})_1'/(Z_{e2})_1']. \tag{3.18}$$

Applying $\psi_3=e^{-iKl}\psi_4$ we find

$$\begin{aligned}&A_0(S_{e0})_3(Z_{e0})_3+A_2(S_{e2})_3(Z_{e2})_3 \\ &\quad+B_0(S_{00})_3(Z_{00})_3+B_1(S_{01})_3(Z_{01})_3 \\ &\quad=e^{-iKl}[A_0(S_{e0})_4(Z_{e0})_4+A_2(S_{e2})_4(Z_{e2})_4 \\ &\quad+B_0(S_{00})_4(Z_{00})_4+B_1(S_{01})_3(Z_{01})_3].\end{aligned} \tag{3.19}$$

Equation (3.19) reduces to

$$A_0\left[(S_{e0})_3(Z_{e0})_3-(S_{e2})_3(Z_{e2})_3\frac{(Z_{e0})_1'}{(Z_{e2})_1'}\right](1-e^{-iKl}) +B_1[(S_{01})_3(Z_{01})_3](1+e^{-iKl})=0. \tag{3.20}$$

Applying $\partial\psi_3/\partial u=-e^{-iKl}\partial\psi_4/\partial u$, we find

$$A_0\left[(S_{e0})_3(Z_{e0})_3'-(S_{e2})_3(Z_{e2})_3'\frac{(Z_{e0})_1'}{(Z_{e2})_1'}\right](1+e^{-iKl}) +B_1[(S_{01})_3(Z_{01})_3'](1-e^{-iKl})=0. \tag{3.21}$$

B_0 does not appear in any of the preceding equations because

$$(S_{00})_1=(S_{00})_2=(S_{00})_3=(S_{00})_4=0. \tag{3.22}$$

However, we can show that $B_0=0$ by equating E_y^* (or H_y) at point 3 and point 4. For this we have

$$\left(\frac{\partial u}{\partial x}\right)_3=\left(\frac{\partial u}{\partial x}\right)_4=0 \quad \text{and} \quad \left(\frac{\partial\gamma}{\partial x}\right)_3=-\left(\frac{\partial\gamma}{\partial x}\right)_4. \tag{3.23}$$

Therefore, we get the additional equation for matching boundary conditions

$$\frac{\partial\psi_3}{\partial\gamma}=-e^{-iKl}\left(\frac{\partial\psi_4}{\partial\gamma}\right). \tag{3.24}$$

Applying this condition, we get

$$B_0[(S_{00})_3(Z_{00})_3](1+e^{-iKl})=0. \tag{3.25}$$

Since Eq. (3.25) must hold for all values of C_0k and ϵ we must have

$$B_0=0. \tag{3.26}$$

Similarly, if we had included B_2 we could show that $B_2=0$ by making use of the boundary condition for E_z^* (or H_z) at points 1 and 2. The above indicates that for the lattice medium of Fig. 8 Eq. (3.5) should contain only odd subscripted B's (but all odd and even subscripted A's) and that the number of unknown $A-B$ constants in ψ should equal the number of points at which tangential boundary conditions are matched. This will still result in one arbitrary unevaluated constant since, as seen from Eqs. (3.20) and (3.21), in order to avoid trivial algebraic solutions the determinant of these two equations must be set equal to zero. This process, while it results in setting up the functional transcendental relationship between k and K, also has the effect of depressing by one the number of simultaneous equations available and thus giving solutions for B_1 and A_2 in terms of A_0.

We define

$$M=(S_{e0})_3(Z_{e0})_3'-\frac{(Z_{e0})_1'}{(Z_{e2})_1'}(S_{e2})_3(Z_{e2})_3',$$
$$N=(S_{01})_3(Z_{01})_3',$$
$$P=(S_{e0})_3(Z_{e0})_3-\frac{(Z_{e0})'}{(Z_{e2})_1'}(S_{e2})_3(Z_{e2})_3, \tag{3.27}$$
$$Q=(S_{01})_3(Z_{01})_3.$$

Now if we set the determinant of Eq. (3.20) and (3.21) equal to zero, we secure a specific example of the general relationship of Eq. (2.15) as applied to the elliptic cylinder lattice medium of Fig. 8:

$$\cot^2\left(K\frac{l}{2}\right)=-\frac{PN}{MQ}. \tag{3.28}$$

The poles and zeros of $-PN/MQ$ (as well as the condition that $-PN/MQ$ must be positive in order to secure K real) are what determine the transmission modes K_T discussed with Eq. (2.15). The functional form of PN/MQ is dependent on the geometry of the obstacle and the geometric mode representation of the wave function made necessary by this geometry.

The accuracy with which PN/MQ represents the physical situation will depend on the number of arbitrary $A-B$ constants in the wave function, Eq. (3.5), and thus on the number of pairs of points at which the cell boundary conditions are matched. The more $A-B$ constants used, the more poles and zeros will occur in Eq. (3.28), and the more closely will they be spaced. This will give more but narrower transmission region representations. As the transmission mode number increases the narrower will be the frequency region covered by each mode.

For a given accuracy of representation, the larger the obstacle cross section in comparison to the unit cell area, the greater the number of $A-B$ constants required.

For a given cell geometry, the higher the frequency, the higher the number of transmission modes which are capable of being supported by the given lattice medium.

Remembering $\pi_z=\psi e^{i(ky-\omega t)}$, we may relate the impedance concepts of Sec. II to the elliptic cylinder medium and rewrite Eq. (3.13):

$$\psi=A_0\left[S_{e0}Z_{e0}-\left(\frac{[Z_{e0}]_1'}{[Z_{e2}]_1'}\right)\times S_{e2}Z_{e2}+i\left(\frac{MP}{NQ}\right)S_{01}Z_{01}\right]. \tag{3.29}$$

3. Propagation function—x direction.—Now let us consider that in Fig. 8 the desired direction of propagation is in the positive x direction. Therefore Eq. (3.2) becomes

$$\pi_z^*=e^{i(Kx-\omega t)}\psi^*. \tag{3.30}$$

The form of ψ^*, Eq. (3.5), remains the same. Equations (3.6), (3.7), and (3.9) remain the same. Equation (3.8) becomes

$$E_\epsilon^*=\frac{-i\omega\mu}{C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}}\times\left[e^{iKx}\frac{\partial\psi^*}{\partial\gamma}-(iKC_0\cosh u\sin\gamma)\pi_z^*\right],$$
$$E_\eta^*=\frac{-i\omega\mu}{C_0(\cosh^2u-\cos^2\gamma)^{\frac{1}{2}}}\times\left[e^{iKx}\frac{\partial\psi^*}{\partial u}+(iKC_0\sinh u\cos\gamma)\pi_z^*\right]. \tag{3.31}$$

Equation (3.10) still holds, and Eq. (3.11) becomes

$$\psi_1^*=(e^{-iKm})\psi_2^*;\quad \frac{\partial\psi_1^*}{\partial u}=-e^{iKm}\frac{\partial\psi_2^*}{\partial u},$$
$$\psi_3^*=\psi_4^*;\qquad \frac{\partial\psi_3^*}{\partial u}=-\frac{\partial\psi_4^*}{\partial u}. \tag{3.32}$$

Applying $\psi_3^*=\psi_4^*$, we get

$$B_1=-B_1=0. \tag{3.33}$$

Applying $\partial\psi_3^*/\partial u=-\partial\psi_4^*/\partial u$ we get

$$A_2=-A_0\frac{(S_{e0})_3(Z_{e0})_3'}{(S_{e2})_3(Z_{e2})_3'}. \tag{3.34}$$

Applying $\psi_1^*=e^{-iKm}\psi_2^*$ we get

$$A_0\left[(Z_{e0})_1-\frac{(S_{e0})_3(Z_{e0})_3'}{(S_{e2})_3(Z_{e2})_3'}(Z_{e2})_1\right](1-e^{-iKm})+A_1[(Z_{e1})_1](1+e^{-iKm})=0. \tag{3.35}$$

Applying $\partial\psi_1^*/\partial u=-e^{-iKm}\partial\psi_2^*/\partial u$ we get

$$A_0\left[(Z_{e0})_1'-\frac{(S_{e0})_3(Z_{e0})_3'}{(S_{e2})_3(Z_{e2})_3'}(Z_{e2})_1\right](1+e^{-iKm})+A_1[(Z_{e1})_1'](1-e^{-iKm})=0, \tag{3.36}$$

$$S=(Z_{e0})'-\frac{(S_{e0})_3(Z_{e0})_3'}{(S_{e2})_3(Z_{e2})_3'}(Z_{e2})_1',$$
$$T=(Z_{e1})_1',$$
$$U=(Z_{e0})_1-\frac{(S_{e0})_3(Z_{e0})_3'}{(S_{e2})_3(Z_{e2})_3'}(Z_{e2})_1, \tag{3.37}$$
$$V=(Z_{e1})_1.$$

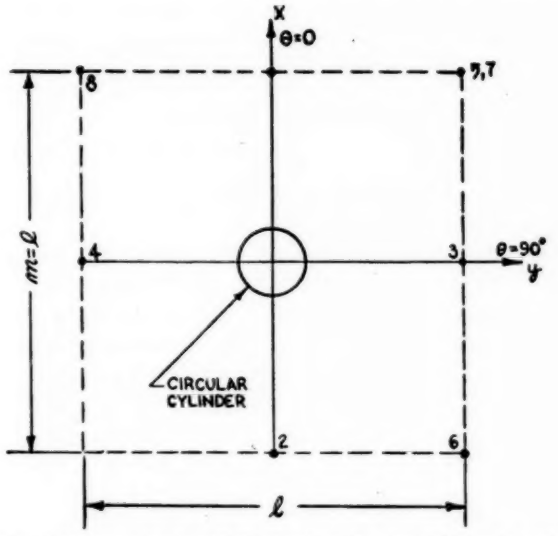


FIG. 9. Diagram of unit cell illustrating, for simplified case, points on cell boundaries where boundary conditions are matched.

Therefore,

$$\cot^2\left(K\frac{m}{2}\right)=-\frac{UT}{SV}. \tag{3.38}$$

By making use of the asymptotic expressions of Eqs. (3.1), it can be shown that when the ellipse is degenerated into a cylinder and the unit cell is made square ($l=m$), then Eqs. (3.28) and (3.38) become equal and reduce to equation derived from consideration of a square array of circular cylinders[43,51] to start with:

$$\cot^2\left(K\frac{l}{2}\right)=-\frac{1}{2}\frac{Z_1'}{Z_1}\left[\frac{Z_0}{Z_0'}+\frac{Z_2}{Z_2'}\right], \tag{3.39}$$

where the combinations of Bessel functions represented by the Z's are all evaluated at $r=l/2$. It should be remembered that for the circular cylinder $Z_{0m}=Z_{em}=Z_m$, and $S_{0m}(c_0k,\ \cos\gamma)\rightarrow\sin m\theta$, and $S_{em}(c_0k,\ \cos\gamma)\rightarrow\cos m\theta$.

4. Higher approximations.—Let us investigate a little more closely the mechanics of evaluating boundary conditions and determining the transcendental relationships between K and k when more than the two pairs of points 1, 2 and 3, 4 are used. Let us refer to Fig. 9. For simplicity of manipulation we shall use circular cylinders arranged in a square array.

For the above situation the boundary conditions for matching pairs of points along the side faces (e.g. points 1 and 2, 5 and 6) are

$$\psi_n{}^*=\psi_{n+1}{}^*,$$

$$\left[\frac{\sin\theta_n}{r_n}\frac{\partial\psi^*}{\partial\theta}-k\cos\theta_n\frac{\partial\psi^*}{\partial(kr)}\right]_n \tag{3.40}$$

$$=\left[\frac{\sin\theta_{(n+1)}}{r_n}\frac{\partial\psi^*}{\partial\theta}-[k\cos\theta_{(n+1)}]\frac{\partial\psi^*}{\partial(kr)}\right]_{(n+1)}.$$

The boundary conditions for the end faces are

$$\psi_{n+2}{}^*=(\psi_{(n+3)}{}^*)e^{-iKl},$$

$$\left[\frac{\cos\theta_{(n+2)}}{r_{n+2}}\frac{\partial\psi^*}{\partial\theta}+k\sin\theta_{(n+2)}\frac{\partial\psi^*}{\partial(kr)}\right]_{(n+2)} \tag{3.41}$$

$$=e^{-iKl}\left[\frac{\cos\theta_{(n+3)}}{r_{(n+3)}}\frac{\partial\psi^*}{\partial\theta}+k\sin\theta_{(n+3)}\frac{\partial\psi^*}{\partial(kr)}\right]_{(n+3)},$$

where $\theta_n=180°-\theta_{(n+1)}$ and $\theta_{(n+2)}=180°-\theta_{(n+3)}$ and $r_n=r_{(n+1)}$.

Evaluation of the boundary conditions Eqs. (3.40) and (3.41) gives the following eight equations with which to evaluate the $A-B$ constants:

$$\begin{aligned}
&A_1(Z_1{}^*)_1+A_3(Z_3{}^*)_1=0,\\
&A_0(Z_0{}^*)_1'+A_2(Z_2{}^*)'+A_4(Z_4{}^*)_1'=0,\\
&\frac{1}{\sqrt2}A_1(Z_1{}^*)_5-\frac{1}{\sqrt2}A_3(Z_3{}^*)_5+B_2(Z_2{}^*)_5=0,\\
&\frac{1}{l}\left[2A_2(Z_2{}^*)_5-B_1(Z_1{}^*)_5\frac{1}{\sqrt2}+3B_3(Z_3{}^*)_5\frac{1}{\sqrt2}\right]\\
&\quad+\frac{k}{\sqrt2}\left[A_0(Z_0{}^*)_5'-A_4(Z_4{}^*)_5'\right.\\
&\quad\left.+B_1(Z_1{}^*)_5'\frac{1}{\sqrt2}+B_3(Z_3{}^*)_5'\frac{1}{\sqrt2}\right]=0,\\
&(1-e^{-iKl})[A_0(Z_0{}^*)_3-A_2(Z_2{}^*)_3+A_4(Z_4{}^*)_3]\\
&\quad+(1+e^{-iKl})[B_1(Z_1{}^*)_3-B_3(Z_3{}^*)_3]=0,\\
&(1-e^{-iKl})[B_1(Z_1{}^*)_3'-B_3(Z_3{}^*)_3']\\
&\quad+(1+e^{-iKl})[A_0(Z_0{}^*)_3'\\
&\quad-A_2(Z_2{}^*)_3'+A_4(Z_4{}^*)_3']=0,\\
&(1-e^{-iKl})\left[A_0(Z_0{}^*)_7+A_1(Z_1{}^*)_7\frac{1}{\sqrt2}\right.\\
&\quad\left.-A_3(Z_3{}^*)_7\frac{1}{\sqrt2}-A_4(Z_4{}^*)_7\right]\\
&\quad+(1+e^{-iKl})\left[B_1(Z_1{}^*)_7\frac{1}{\sqrt2}\right.\\
&\quad\left.+B_2(Z_2{}^*)_7+B_3(Z_3{}^*)_7\frac{1}{\sqrt2}\right]=0,\\
&(1-e^{-iKl})\left\{\frac{1}{l}\left[B_1(Z_1{}^*)_7\frac{1}{\sqrt2}-3B_3(Z_3{}^*)_7\frac{1}{\sqrt2}\right]\right.\\
&\quad\left.+\frac{k}{\sqrt2}\left[B_1(Z_1{}^*)_7'\frac{1}{\sqrt2}+B_2(Z_2{}^*)_7'\right.\right.
\end{aligned} \tag{3.42}$$

$$+B_3(Z_3^*)_7'\frac{1}{\sqrt{2}}\Big]\Big\}+(1+e^{-iKl})$$

$$\times\Big\{\frac{1}{l}\Big[-A_1(Z_1^*)_7\frac{1}{\sqrt{2}}-2A_2(Z_2^*)_7$$

$$-3A_3(Z_3^*)_7\frac{1}{\sqrt{2}}\Big]+\frac{k}{\sqrt{2}}\Big[A_0(Z_0^*)_7'$$

$$+A_1(Z_1^*)_7'\frac{1}{\sqrt{2}}-A_3(Z_3^*)_7'$$

$$\times\frac{1}{\sqrt{2}}-A_4(Z_4^*)_7'\Big]\Big\}=0.$$

To simplify computations somewhat, let us replace our square unit cell by a circularly symmetric one of equal area, i.e.,

$$\beta^2=\pi\beta^2 \quad \text{or} \quad \beta=l/(\pi)^{\frac{1}{2}}. \tag{3.43}$$

This substitution results in the following relationships:

$$r_1=r_2=-r_3=r_4=r_5=r_6=r_7=r_8=l/(\pi)^{\frac{1}{2}}. \tag{3.44}$$

The θ-values remain unchanged. Therefore, we can drop the point evaluation notation in Eqs. (3.42) for the Z^* functions. All Z^* function evaluations are done at $r=l/(\pi)^{\frac{1}{2}}$. The worst errors introduced by this approximation are for the corner points where r should be $l/\sqrt{2}$ instead of $l/(\pi)^{\frac{1}{2}}$. At points 1, 2, 3, 4 the error is in the other direction; here, r should equal $l/2$, instead of $l/(\pi)^{\frac{1}{2}}$.

Using the first four of Eqs. (3.42) we can eliminate A_2, A_3; B_2; and B_3:

$$B_2=-2Z_1^*A_1/Z_2^*,$$

$$A_3=-Z_1^*A_1/Z_3^*,$$

$$A_2=-\frac{A_0(Z_0^*)'+A_4(Z_4^*)'}{Z_2^{*\prime}},$$

$$B_3=-\frac{\left(k\beta\left(\frac{\pi}{2}\right)^{\frac{1}{2}}(Z_0^*)'-2\frac{(Z_0^*)'}{(Z_2^*)'}Z_2^*\right.}{3Z_3^*+k\beta\left(\frac{\pi}{2}\right)^{\frac{1}{2}}Z_3^{*\prime}}A_0$$

$$+\frac{\left[2\frac{(Z_4^*)'}{(Z_2^*)'}Z_2^*+k\beta\left(\frac{\pi}{2}\right)^{\frac{1}{2}}(Z_4^*)'\right]}{3Z_3^*+k\beta\left(\frac{\pi}{2}\right)^{\frac{1}{2}}(Z_3^*)'}A_4 \tag{3.45}$$

$$-\frac{\left(\frac{k\beta(Z_1^*)'}{2}\right)-\frac{Z_1^*}{\sqrt{2}}}{3Z_3^*+\left[k\beta\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\right](Z_3^*)'}B_1.$$

By substituting these values in the last four of Eqs. (3.42) we get the transcendental relationship between k and K as the solution of a quartic equation.

$$E\cot^4(K\beta)+F\cot^3(K\beta)+G\cot^2(K\beta)$$
$$+H\cot(K\beta)+L=0. \tag{3.46}$$

In Eq. (3.46) E, F, G, H, and L, are all complicated polynomials of Z^* and Z^*.

Equation (3.46) symbolizes the solution for 4 pairs of matching points whereas Eq. (3.39) represents the solution for the minimum of 2 pairs of matching points. We will now study the solution for the intermediate case of 3 pairs of matching points by eliminating points 3 and 4 in Fig. 9. We shall continue to use the approximation of Eq. (3.44). The equation for ψ^* becomes

$$\psi^*=A_0Z_0^*+A_1Z_1^*\cos\theta+A_2Z_2^*\cos2\theta$$
$$+A_3Z_3^*\cos3\theta+B_1Z_1^*\sin\theta+B_2Z_2^*\sin2\theta, \tag{3.47}$$

and Eqs. (3.42) reduce to

$$A_1Z_1^*+A_3Z_3^*=0,$$

$$A_0(Z_0^*)'+A_2(Z_2^*)'=0,$$

$$A_1\frac{Z_1^*}{\sqrt{2}}-A_3\frac{Z_3^*}{\sqrt{2}}+B_2Z_2^*=0,$$

$$A_22Z_2^*-B_1\frac{Z_1^*}{\sqrt{2}}$$

$$+\sqrt{2}k\beta\left[A_0(Z_0^*)'+B_1\frac{(Z_1^*)'}{\sqrt{2}}\right]=0,$$

$$\left[A_0Z_0^*+A_1\frac{Z_1^*}{\sqrt{2}}-A_3\frac{Z_3^*}{\sqrt{2}}\right]$$

$$-i\cot(K\beta)\left[B_1\frac{Z_1^*}{\sqrt{2}}+B_2Z_2^*\right]=0, \tag{3.48}$$

$$\left[B_1\frac{Z_1^*}{\sqrt{2}}+\sqrt{2}k\beta\left[B_1\frac{(Z_1^*)_2'}{\sqrt{2}}+B_2(Z_2^*)'\right]\right]$$

$$-i\cot(K\beta)\Big\{-A_1\frac{Z_1^*}{\sqrt{2}}$$

$$-2A_2Z_2^*-3A_3\frac{Z_3^*}{\sqrt{2}}+\sqrt{2}k\beta$$

$$\times\left[A_0(Z_0^*)'+A_1\frac{(Z_1^*)'}{\sqrt{2}}-A_3\frac{(Z_3^*)'}{\sqrt{2}}\right]\Big\}=0.$$

The solution of Eqs. (3.48) leads to a quadratic equation

$$A\cot^2(K\beta)+B\cot(K\beta)+C=0, \tag{3.49}$$

where

$$
\begin{aligned}
A=&\,4\sqrt{2}k\beta(Z_0^*)'\frac{(Z_1^*)^2}{2}-\sqrt{2}k\beta(Z_0^*)'\frac{Z_1^*}{\sqrt{2}}[k\beta(Z_1^*)']\\
&-6\frac{Z_2^*}{(Z_2^*)'}(Z_0^*)'\left(\frac{Z_1^*}{\sqrt{2}}\right)(k\beta(Z_1^*)')\\
&+\left[k\beta\frac{(Z_3^*)'}{Z_3^*}Z_1^*\right]\left[\frac{Z_1^*}{\sqrt{2}}\right]\\
&\times\left[\sqrt{2}k\beta(Z_0^*)'-2\frac{Z_2^*}{(Z_2^*)'}(Z_0^*)'\right],\\
B=&\left[4(k\beta)^2(Z_0^*)'(Z_1^*)'-4\frac{Z_2^*}{(Z_2^*)'}(Z_0^*)'Z_1^*\right]\\
&\times(iZ_1^*)+\left[-\frac{Z_1^*}{\sqrt{2}}+k\beta(Z_1^*)'\right]\\
&\times\left[k\beta(Z_1^*)'+\sqrt{2}Z_1^*+k\beta\frac{(Z_3^*)'}{Z_3^*}Z_1^*\right](iZ_0^*)\\
&+\left[-2\frac{Z_2^*}{(Z_2^*)'}(Z_0^*)'+\sqrt{2}k\beta(Z_0^*)'\right]\\
&\times\left[2k\beta\frac{(Z_2^*)'}{Z_2^*}Z_1^*\right]\left(i\frac{Z_1^*}{\sqrt{2}}\right),\\
C=&\left[\sqrt{2}k\beta(Z_0^*)'-2\frac{Z_2^*}{(Z_2^*)'}(Z_0^*)'\right]\\
&\times\left[\frac{Z_1^*}{\sqrt{2}}+k\beta(Z_1^*)'\right](\sqrt{2}Z_1^*)\\
&-(Z_0^*)\left[k\beta(Z_1^*)'-\frac{Z_1^*}{\sqrt{2}}\right]\left[2k\beta\frac{(Z_2^*)'}{Z_2^*}Z_1^*\right].
\end{aligned}
\tag{3.50}
$$

Thus we see that the relatively simple form of Eqs. (3.28), (3.38), and (3.39) are the result of the choosing of the very special points 1, 2 and 3, 4 at which to evaluate the unit cell boundary conditions. Of course, the simplicity has its penalty in that the polynomial functions of k on the right-hand sides of these equations do not contain as many poles and zeros for a given interval of k as do the more complicated solutions provided by Eqs. (3.46) and (3.49). Therefore, Eq. (3.49) yields a relationship between K and k which contains more pass and stop bands than does Eq. (3.39), and similarly Eq. (3.46) is more accurate than Eq. (3.49).

The need for greater number of matching points becomes more urgent as the obstacle dimensions become larger compared to the unit cell dimensions and/or as the free-space wavelength of the electromagnetic radiation passing through the lattice becomes smaller and smaller compared to the unit cell dimensions (especially the transverse dimensions). These requirements provide the connection between the geometric modes used in the wave function ψ and the transmission modes resulting from the pass and stop bands inherent in the transcendental relationship between K and k. The larger the obstacle, the more terms are needed in the series solution of ψ to allow the diffraction solution appropriate to the geometry of the obstacle to be fitted to the rectangular (or square) cell boundary. This requires more pairs of matching points and simultaneously provides a more accurate representation of the higher transmission modes.

5. *Numerical calculations.*—Figure 4, as drawn, is simply a generalized sketch or picture of the behavior of K_T as a function of k and the geometry of the obstacle. Equations (3.28), (3.38), (3.39), (3.46), and (3.49) are the basic transcendental equations governing the relationship between K, the average lattice wave number, and k, the free-space wave number. In addition, of course, these equations contain the geometric parameters which describe the spacing of the obstacles and the size of the obstacles. These two parameters occur only in a multiplicative relationship with K and k. In particular, only the spacing factor is multiplied with K, but both the spacing and size factors appear as multiplying factors of k. This latter fact allows us to always normalize the geometric constants of the lattice in terms of the free-space wavelength λ. Such normalization is used in the numerical calculations since it permits one set of data to furnish information concerning the behavior of a given physical lattice medium as the frequency of the electromagnetic radiation is varied; or alternatively, for a given frequency, it gives information about the behavior of different physical lattice media. In addition to spacing and size of the obstacles of a two-dimensional lattice, the geometric shape of the obstacle is important in determining the lattice transmission behavior. Inspection of the above referenced equations indicate that the obstacle shape factor also enters into these equations. The functional form changes as the geometric shape changes. Thus, when considering circular cylinders we find that Bessel functions are involved; whereas, when elliptic cylinders are considered, these are replaced by the more general Mathieu functions.[41]

Since the primary function of this paper was not the establishment of numerical function tables, a compromise had to be struck between calculation time involved and usefulness of the data derived. To perform calculations of reference tables on Eqs. (3.28), (3.38), (which involve Mathieu functions) or (3.46), (3.49) (which are very complicated algebraically), one should have available a large machine computer. The illustrative calculations for this paper were performed on Eq. (3.39). This equation illustrates the transmission behavior of lattice media as the various critical parameters are varied and yet permits calculations to be accomplished with

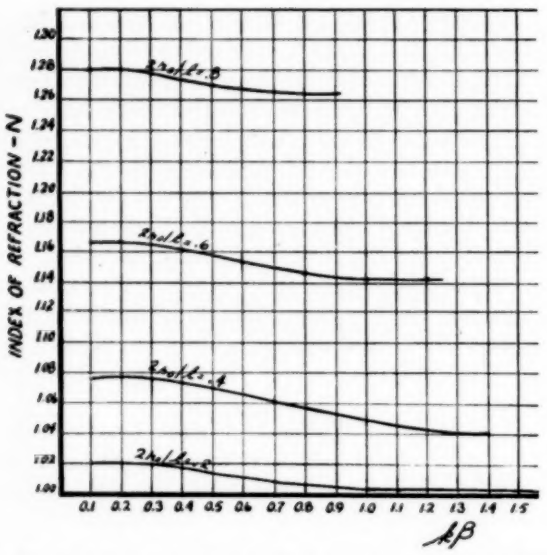


FIG. 10. Index of refraction in first pass band as a function of cylinder size, spacing and frequency for the E plane case.

reasonable effort without the use of a large machine computer.

For the E plane case the lattice medium simulates a solid dielectric which delays the electromagnetic waves passing through it compared with its phase velocity in free space. At values of λ large compared with cell and obstacle size, the simulation is very good. As λ gets smaller we come to a stopping band. In this first low pass band the index of refraction is fairly constant (as it should be for a true dielectric). Towards the upper part of this band, though, the index of refraction starts to vary with frequency which indicates the dispersive character of the medium. In the second pass band the medium would exhibit even greater dispersive tendencies. Also the second band is narrower than the first, the third pass band still narrower and so on. Because only two pairs of matching points (points 1, 2, 3, and 4 in Fig. 8) were used in the derivation of Eq. (3.39) the higher pass bands are inadequately represented. The greater the number of matching points the more accurately is any single pass band represented and the greater the number of pass bands which are adequately represented.

In Fig. 10 is plotted the index of refraction *versus* $k\beta$ for the first pass band. Since, $k\beta=(2\pi f/V)\beta$ and if β is kept constant, then $k\beta$ is proportional to frequency. In order to illustrate the effect of change of obstacle size, the functions were evaluated for four different ratios of obstacle to cell size. r_0 is the obstacle radius and l is the cell length. Since the greater values of N mean greater phase velocity delays, it can be seen from Fig. 10 that the larger the obstacle the greater the phase velocity delay. Also illustrated by Fig. 10 is the dispersive characteristic mentioned above. The smaller the obstacle, the less the variation of N during the low frequency portion of the pass band. Since Fig. 10 is plotted only for the first pass band the narrowing of succeeding higher pass bands is not illustrated. However, the narrowing of the first pass band as the obstacle size increases is illustrated by the shorter lengths of the curve as the obstacle size increases.

In order to illustrate the behavior of the higher pass bands Fig. 11 and 12 were plotted. These show $\cot K\beta$ *versus* $k\beta$ for a constant ratio of $2r_0/l$. $\cot K\beta$ was chosen as the ordinate rather than $K\beta$ in order to eliminate the question of multi-values. Another way to handle this question would have been to plot $K\beta$ *vs* $k\beta$ in a manner similar to Fig. 4. However, for purposes of illustrating the behavior of the successive pass bands the plots of Figs. 11 and 12 are simpler. The plots illustrate that for a given ratio $2r_0/l$ the higher pass bands get narrower and narrower. As the ratio $2r_0/l$ gets higher, corresponding pass bands get narrower. Thus, for $2r_0/l=0.1$ the first pass band extends from $k\beta=0$ to $k\beta=1.44$ while for $2r_0/l=0.5$ the first pass band extends from $k\beta=0$ to $k\beta=1.32$.

Another point illustrated by Figs. 11 and 12 is the limitation of the function Eq. (3.39) which is the result of the minimum number of points at which the boundary conditions were matched. As the obstacle gets larger this approximation becomes progressively worse. Thus, where for $2r_0/l=0.1$ we are able to plot two complete pass bands and most of the third one, when $2r_0/l=0.5$, the Eq. (3.39) yields only one pass band and a portion of the second.

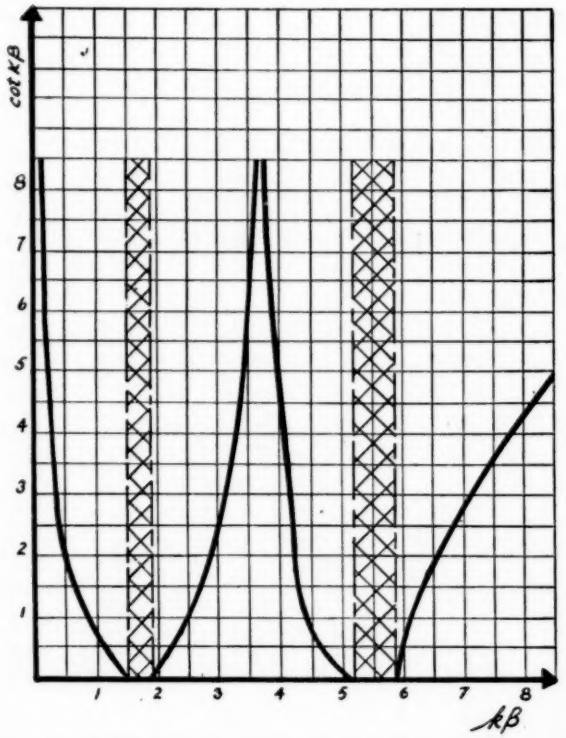


FIG. 11. Relationship between average lattice propagation constant K and frequency of the propagated energy for E plane case when $2r_0/\lambda=0.1$ (cross-hatched areas are stop bands).

Figures 13 and 14 are the same as Figs. 11 and 12 except for the H plane case instead of the E plane. For H plane polarization the lattice medium simulates a parallel plate wave guide. The parallel plate wave guide increases the phase velocity of electromagnetic energy passing through it as compared with the phase velocity in free space.

For the H plane case there is a lower limit (other than zero) to the frequency of the electromagnetic energy which will be transmitted by the medium. Above this cut-off frequency the medium has successive pass and stop bands just as was exhibited by the delay medium (E plane case). The first stop band is seen to be a function of obstacle size. Thus, when $2r_0/l=0.1$ the first

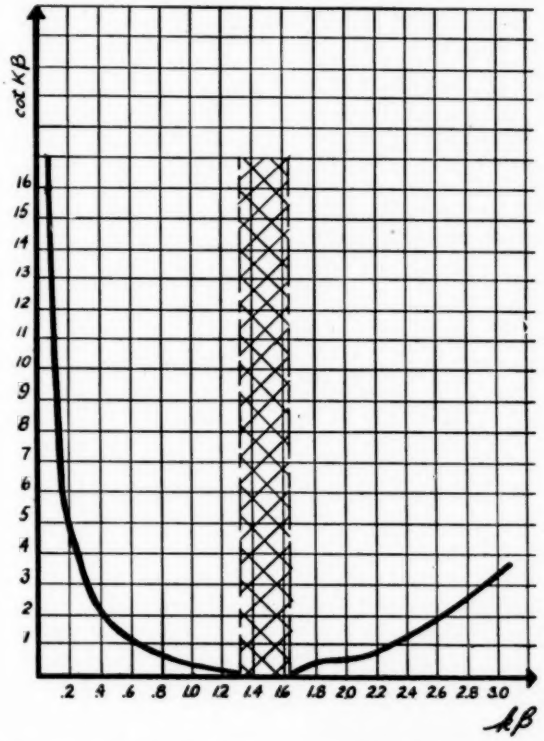


FIG. 12. Relationship between average lattice propagation constant K and frequency of the propagated energy for E plane case when $2r_0/\lambda=0.5$ (cross-hatched area is stop band).

stop band extends from $k\beta=0$ to $k\beta=1$; whereas, for $2r_0/l=0.5$ the first stop band extends from $k\beta=0$ to $k\beta=2.7$.

Figures 13 and 14 also indicate the limitations of the simple approximation of the boundary conditions. For $2r_0/l=0.1$ we can plot the first 3 pass bands (albeit not with 100 percent accuracy), while for $2r_0/l=0.5$ Eq. (3.39) yields only two pass bands before "blowing up" completely.

The general tendency for narrowing the pass bands as the frequency goes up and the narrowing or corresponding pass bands as the obstacle size increases is also exhibited by Figs. 13 and 14 although not as clearly as in the preceding case. This is due to the fact that for the

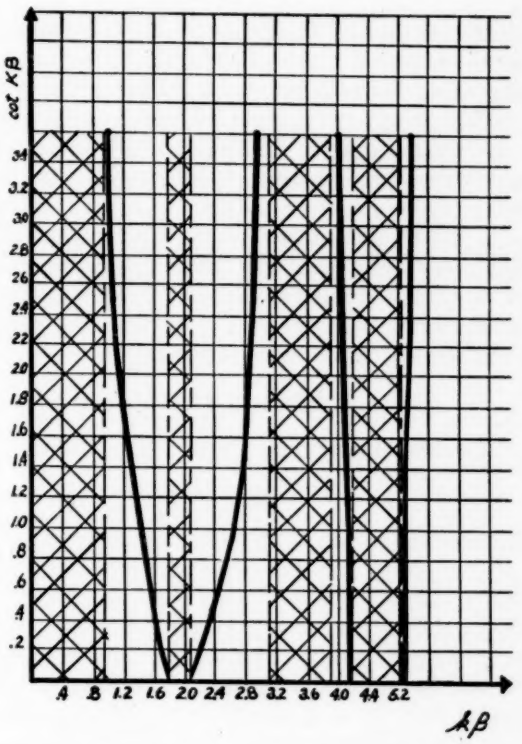


FIG. 13. Relationship between average lattice propagation constant K and frequency of the propagated energy for H plane case when $2r_0/\lambda=0.1$ (cross-hatched areas are stop bands).

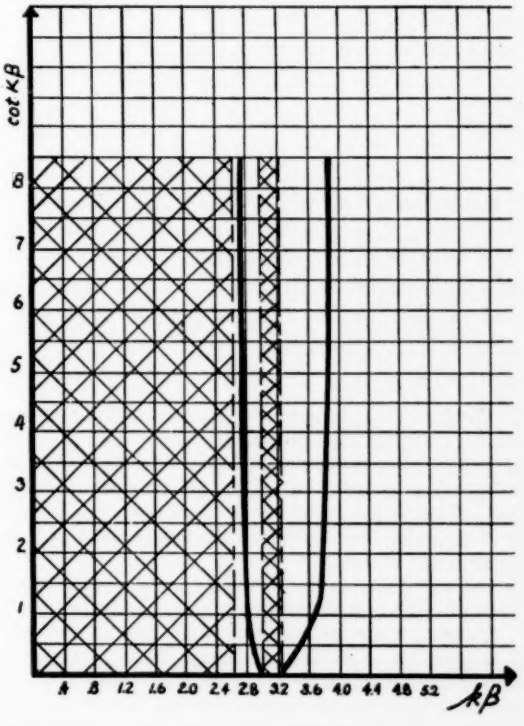


FIG. 14. Relationship between average lattice propagation constant K and frequency of the propagated energy for H plane case when $2r_0/\lambda=0.5$ (cross-hatched areas are stop bands).

H plane case the stop bands of Eq. (3.39) are caused by either the negative values of $\cot^2 K\beta$ going to infinity or staying finite. For example, in Fig. 13 the first stop band has $\cot^2 K\beta$ going to minus infinity, the second stop band has only finite negative values of $\cot^2 K\beta$, the third an infinite value etc.; whereas, in Fig. 11 all stop bands maintain finite negative values for $\cot^2 K\beta$. Thus in Figs. 13 and 14, if one separates the stop bands in which $\cot^2 K\beta$ goes to minus infinity from those where $\cot^2 K\beta$ stays finite and negative, each set independently exhibits the same behavior with variation in frequency and obstacle size as does the E plane case.

SECTION IV

Conclusions and Practical Applications

1. Conclusion

The preceding material has attempted to carry out the objectives of this paper which were outlined in Section I. We have investigated the behavior of two-dimensional lattice media and structures in the presence of electromagnetic energy. The various factors governing the transmission of electromagnetic energy by these structures have been discussed, a unifying impedance concept has been set forth and basic mathematical functions have been developed for the elliptic cylinder case which can serve to approximate almost any two-dimensional lattice medium.

The impedance concept and basic functions are not limited by the usual approximations which have been applied heretofore, namely, that the wavelength must be much larger than obstacle and cell size and that the obstacle size must be very much smaller than the cell size.

2. Practical Applications

If the functions developed herein are properly tabulated by means of large mechanical computers the uses of the two-dimensional lattice structure would be greatly extended and facilitated. Following is a partial list of applications in which greater design facility would result from the proper tabulation of the functions developed herein:

(a) Design of electromagnetic wave focusing devices—including all delay, variable N and wave-guide lenses.

(b) Design of prescribed E and H distributions over lens antenna apertures in order to achieve special main and side lobe relations and shapes.

(c) Design of rapid scan lens antennas by rotation of individual lattice structure obstacles.

(d) Design of the lattice structure as a circuit component (wave filter type) of microwave transmission circuits.

(e) Design of traveling wave tube amplifiers and oscillators.

(f) Design of a variable polarization device. By varying the orientation of the obstacles and the total length of the lattice structure one varies the relationship between K_T and K_T^*, and one can change the polarization of transmitted electromagnetic energy.

(g) There are many applications where it is desired to maximize the reflection of electromagnetic energy at a boundary of a lattice structure. While the cellular method of analysis is not specifically designed to permit this, it furnishes the basis for an intelligent design estimate by defining the boundaries of the stop or non-transmission bands.

3. Suggestions for Further Research

The following are problems worthy of further effort in this field:

(a) Machine computation and tabulation of the lattice functions developed herein.

(b) Refinement of the approximations used in Sec. II in applying the impedance concept to the transition region of a lattice structure.

(c) Analysis of lattice media and structures made up of finitely conducting obstacles.

(d) Quantitative study of the effect of lattice structure aperture size on the E and H distribution over input and output faces.

(e) Application of the concepts of Secs. II and III to three-dimensional lattice structures (i.e., periodic repetition in all three dimensions as might be obtained with ellipsoidal obstacles).

(f) Application of other than the standard Wigner-Seitz boundary conditions [e.g., the one indicated in Eq. (2.5)]. This would result in a transcendental function relationship between K and k which is different from $\cot^2(K(l/2)) = f(k)$.

(g) Application of integral equation techniques to the boundary value solutions of Sec. III.

APPENDIX

Glossary

A_p—Fourier series coefficient

a_0—Semi-major axis—elliptic cylinder coordinates

B_p—Fourier series coefficient

b_0—Semi-minor axis—elliptic cylinder coordinates

β—$l/\sqrt{\pi}$ = equivalent circular radius of square lattice unit cell

C_m—[see Eq. (3.4)]

$\mathbf{d}_1$—Generalized lattice basis vector

$\mathbf{d}_2$—Generalized lattice basis vector

D_m—[see Eq. (3.4)]

$\bar{E}$—Electric field vector

ϵ—Dielectric constant

ϵ_0—Dielectric constant of free space

ξ—Elliptic cylinder coordinate

η—Elliptic cylinder coordinate

e—Eccentricity—elliptic cylinder coordinate

F_s—Lattice space form factor [see Eq. (2.35)]

G_p—Conductance of equivalent lattice network

$\bar{H}$—Magnetic field vector

H_{sq}—I_q

I_q—Equivalent inegrated current as a result of the magnetic field distribution over the entrance face of the lattice unit cell

i—$-1^{\frac{1}{2}}$

J_n—Bessel function of order η

K_T—Average or over-all lattice propagation constant in the Tth pass band

k—$2\pi/\lambda = \omega/V_0$ = free-space wave number

λ—Free-space wavelength

l—Lattice unit cell dimension in y direction

m—Lattice unit cell dimension in x direction

μ—Permeability
μ_0—Permeability of free space
M—Lattice function [see Eq. (3.27)]
N—Index of refraction of lattice medium; also lattice function [see Eq. (3.27)]
N_n—Neumann function of order n
γ—Elliptic cylinder coordinate
ω—$2\pi f = 2\pi(V_0/\lambda)$
π_z—Hertz vector component in z direction for H plane case; also Bloch wave
π_z^*—Hertz vector component in z direction for E plane case
p—Subscript denoting lattice impedance mode
π_{zq}—$\int\int \pi z \cdot d\bar{s}$ = Bloch wave integrated over cell surface
P—Lattice function [see Eq. (3.27)]
Q—Lattice function [see Eq. (3.27)]
r_0—Radius of circular cylinder lattice obstacle
$\mathbf{r}_{l_1 l_2}$—Generalized lattice vector
$R_{em}^{1,2}$—Even radial Mathieu function of order m, as defined in Stratton[61]
$R_{0m}^{1,2}$—Odd radial Mathieu function of order m, as defined in Stratton[61]
ψ—Bloch wave amplitude factor
$\bar{\psi}$—Complex conjugate of Bloch wave function
S_y—Real part of Poynting vector (in y direction)
S_z—Imaginary or reactive part of Poynting vector (in x direction)
S_{em}—Even angular Mathieu function of order m as defined in Stratton[61]
S_{0m}—Odd angular Mathieu function of order m as defined in Stratton[61]
S—Lattice function [see Eq. (3.37)]
t—Time
θ—Measure of lack of full restoration of the electromagnetic wave at the lattice unit cell exit face as compared with the entrance face [see Eq. (2.51)], also polar coordinate
T—Lattice function [see Eq. (3.37)]
u—Elliptic cylinder coordinate
U—Lattice function [see Eq. (3.37)]
V_0—Velocity of electromagnetic wave in free space
V—Velocity of electromagnetic wave in lattice medium; also lattice function [see Eq. (3.37)]
X—Coordinate axis
Y—Coordinate axis
z—Coordinate axis; direction in which lattice medium is homogeneous
Z_c—Characteristic impedance of equivalent lattice network
Z_{aa}—Mesh impedance of equivalent lattice network
Z_{bb}—Mesh impedance of equivalent lattice network
Z_{ab}—Mutual impedance of equivalent lattice network
Z_{ba}—Mutual impedance of equivalent lattice network
Z_0—Impedance of free space
Z_{TC}—Characteristic impedance of equivalent lattice network in the Tth pass band
Z_s—Equivalent lattice network terminal impedance [see Eq. (2.49)]
Z_{em}—[see Eq. (3.3)]
Z_{0m}—[see Eq. (3.3)]
Z_m—Asymptote of $Z_{em}Z_{0m}$ for circular cylinder case

APPENDIX B. BIBLIOGRAPHY

1. F. Abeles, Ann. Phys. **5**, 596 (1950); **5**, 706 (1950).
2. W. S. Ament, "Preliminary Report on Electromagnetic Wave Propagation in a Composite Medium," The Laboratory, February, 1948.
3. J. Bardeen and W. Shockley, Phys. Rev. **80**, 69 (1950).
4. F. Berz, Proc. Inst. Elec. Engrs. (London) **98**, No. 51, 47 (1951).
5. F. Bloch, Z. Physik **52**, 555 (1928).
6. Grady, Pearson, and Peoples, "Final Report on Contract N17335-11832 on Microwave Optics," Oregon State College.
7. W. H. Bragg and W. L. Bragg, *X-Rays and Crystal Structure*. (George Bell & Sons, London, 1915).
8. H. Bremekamp, Physik **6**, 136 (1926).
9. L. Brillouin, "Wave guides for slow waves," Cruft Laboratory Report No. 4, Harvard University, 1946.
10. L. Brillouin, *Propagation in Periodic Structures* (McGraw-Hill Book Company, Inc., New York, 1946).
11. L. Brillouin, *Structures des corps solides* (Hermann & Cie, Paris, 1937).
12. L. Brillouin Congr. Intern. Electricité (Paris) **2**, 739 (1932).
13. J. Brown, Proc. Inst. Elec. Engrs. (London) **97**, 45 (1950).
14. J. Brown, "The Design of Metallic Delay Dielectrics," G. B. Radio Research Development Establishment, Report No. 345, May, 1949, p. 5.
15. J. Brown and S. S. D. Jones, Electronic Eng. (London) **22**, 127, 183, 227, 264, 358, 429 (1950).
16. C. A. Cochrane, Proc. Inst. Elec. Engrs. (London) **97**, 72 (1950).
17. S. B. Cohn, "A theoretical and experimental study of a waveguide filter structure," Tech. Report No. 39, Cruft Laboratory, Harvard University.
18. S. B. Cohn, J. Appl. Phys. **21**, 674 (1950).
19. S. B. Cohn, "First quarterly report on investigation of metallic delay lenses," Signal Corps Contract No. W36-039-ac-38250 and 38246 with Sperry Gyroscope Company.
20. S. B. Cohn, J. Appl. Phys. **20**, 257 (1949).
21. S. B. Cohn, J. Appl. Phys. **22**, 628 (1951).
22. R. W. Corkum, "Isotropic artificial dielectric," Air Force Cambridge Research Laboratories, March, 1951.
23. H. B. DeVore and H. A. Iams, RCA Rev. **1948**, 721.
24. G. Eckert, Ann. Telecommun. **4**, 142, 223 (1949).
25. G. Estrin, Proc. Inst. Radio Engrs. **39**, 821 (1951).
26. G. Estrin, J. Appl. Phys. **21**, 667 (1950).
27. H. Goldstein, "Theory of corrugated transmission lines and wave guides," Radiation Laboratory Report No. 496, (1944).
28. H. Goldstein, Doctoral thesis, Dept. of Physics, Massachusetts Institute of Technology, 1943).
29. W. W. Hansen and E. L. Chu, J. Appl. Phys. **18**, 996 (1947).
30. A. E. Heins, Proc. Inst. Radio Engrs. **38**, 192 (1950).
31. A. E. Heins, "The reflection of an *EM* plane wave by an infinite set of plates," (1) *H* Plane Case, August, 1948; (2) *E* Plane Case, November, 1948. Cambridge Field Station.
32. S. S. D. Jones and J. Brown, Nature **163**, 324 (1949).
33. W. E. Kock, Proc. Inst. Radio Engrs. **38**, 192 (1950).
34. W. E. Kock, "Metal Plate Lenses for Microwaves," Bell Telephone Laboratory Report No. MM-45-160-23, March 23 (1945).
35. W. E. Kock, Proc. Inst. Radio Engrs. **34**, 828 (1946).
36. W. E. Kock, Bell System Tech. J. **27**, 58 (1948).

36(a). P. E. Krasnushkin, Zhur. Tekh. Fiz. **20**, 1065 (1950).

37. B. A. Lengyel, "Physical optics of metal-plate media; Part I, theoretical considerations," Naval Research Laboratory Report No. 3534.
38. B. A. Lengyel, "Physical optics of metal plate media; Part II, experimental studies," Naval Research Laboratory Report No. 3599, January 3, 1950.
39. Leonard Lewin, J. Inst. Elec. Engrs. (London) **94**, 64 (1947).
40. S. J. Mason, "Metal plate lens formulas," Massachusetts Institute of Technology Radiation Laboratory Report No. 54, May 1, 1945.
41. R. H. McFee, "Design and analysis of a microwave lens antenna," Bumblebee Report, Submarine Signal Company, Boston, Massachusetts.
42. "Dielectric and microwave lenses," Ninth Quarterly Progress Report, Northwestern University.
43. S. Pasternack, "Microwave lenses," Final Report on U. S. Air Force Contract W28-099-ac-112, University of Pennsylvania, July, 1947.
44. Lord J. W. S. Rayleigh, Phil. Mag. **34**, 481 (1892).
45. Lord J. W. S. Rayleigh, Phil. Mag. **24**, 145 (1887).
46. J. Ruze, Proc. Inst. Radio Engrs. **38**, 53 (1950).
47. J. Ruze, "Preliminary report on the experimental determination of the reflection coefficient of a metal-plate interface," Cambridge Field Station, Report No. E4056, April, 1949.
48. F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940).
49. F. Seitz, *Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943).
50. J. C. Slater, Phys. Rev. **45**, 794 (1934).
51. D. Slepian, "Propagation of waves in a lattice of conductors," Project V-7, Research Staff, Harvard University.
52. Otto Stuetzer, Proc. Inst. Radio Engrs. **38**, 1053 (1950).
53. C. Susskind, Wireless World **55**, 370 (1949).
54. C. Susskind, "Investigation of obstacle-type artificial dielectrics," Yale University, Final Report U. S. Air Force Contract, AF 19(122)-270.
55. J. E. Walsh, Proc. Inst. Radio Engrs. **38**, 192 (1950).
56. J. R. Whinnery, "Applications of potential theory to the analysis and synthesis of linear arrays," Hughes Aircraft Company Electronics Dept., Technical Memo No. 199, October 1, 1948.
57. E. P. Wigner and F. Seitz, Phys. Rev. **43**, 804 (1933); **46**, 509 (1934).
58. Wigner, Bouckaert, and Smoluchowski, Phys. Rev. **50**, 58 (1936).
59. G. Wilkes, Proc. Inst. Radio Engrs. **36**, 206 (1948).
60. D. E. Spencer, J. Appl. Phys. **22**, 386 (1951).
61. J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941).

Letters to the Editor

Pure Gliding of Metal Crystals

HIROSHI KANZAKI
Institute of Industrial Science, University of Tokyo, Chiba City, Japan
(Received March 9, 1953)

RÖHM and Kochendörfer[1] have made experiments on the pure gliding of aluminum crystals. Their results show that the values of both the critical shear stress and the coefficient of shear hardening are lowered considerably compared with those obtained in the usual extension experiments. The writer has performed experiments of pure gliding on copper single crystals using improved apparatus. The results obtained will be reported here.

As pointed out by W. T. Read,[2] the method of loading the specimen is rather doubtful in the Röhm's experiments, and the stress distribution is far from pure homogeneous shear.

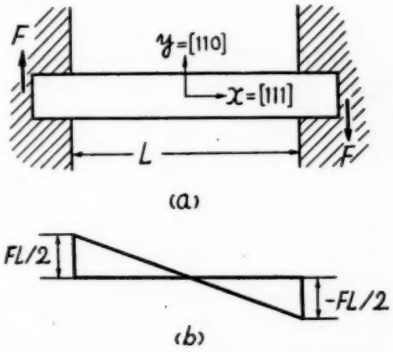


FIG. 1. Schematic diagram of Röhm's experiment. (a) Initial arrangement (from reference 2). (b) Distribution of bending moment along x axis.

Figure 1(a) shows schematically the initial arrangements in the Röhm experiment, where the [111] direction is parallel to the long axis of the specimen and the [110] slip direction is vertical. The distribution of bending moment along the long axis is shown in Fig. 1(b), and it is apparent that the condition of pure gliding is satisfied only at the middle point where the bending moment is zero. This state of affairs should be unavoidable in such types of loading methods.

Considering this point, the writer has measured the strain distribution, not only along x direction but along y direction, and thus separated the two types of strain; these are the shear strain due to shearing force F (which is constant along x) and the bending strain due to bending moment (which varies linearly along x).

As pointed out by Read,[2] the ends are not allowed to move horizontally in the arrangement of Fig. 1 and the unknown axial force is expected to result in the course of deformation. The writer has adopted the cantilever type of loading system, as shown in Fig. 2, in order to allow the horizontal motion of the end.

Specimens of single crystals were prepared from electrolytic copper by the vacuum melting method. They were about 6 mm in diameter and 30 mm in length.

As shown in Fig. 2 the single crystal is settled at the end A and the force F is applied at the free end B along the [110] slip direction of the crystal. Special attention was paid to maintaining the coincidence between the direction of force and the slip direction of the crystal, in response to the rotation due to bending strain

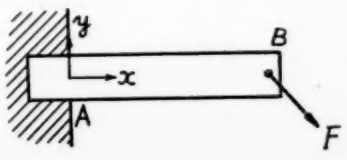


FIG. 2. Improved method of experiments.

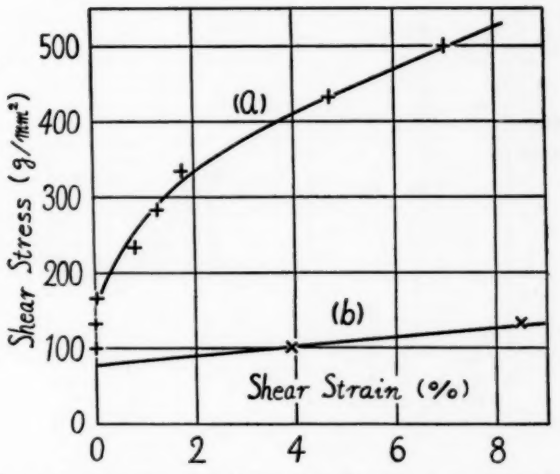


FIG. 3. Shear stress *vs* resolved shear strain for copper crystal. (a) Curve for pure gliding. (b) Curve obtained when all parts of deformation are considered due to shear.

in the course of deformation. The residual strain of the specimen was measured by markings on the crystal surface by means of the optical comparator within the accuracy of 0.001 mm.

Our results on several specimens show that the deformation due to plastic bending is far larger than the deformation due to shearing force F. The local strains along the long axis of specimen change their sign from tension to compression along y axis of Fig. 2, and they decrease with x passing from the fixed end A to the free end B.

Figure 3(a) shows the shear stress *versus* resolved shear strain curve due to shearing force F. This curve was calculated from the data obtained near the free end of the specimen where bending moment takes the lowest value along x axis.

When all parts of the rotational deformation of the specimen are considered due to shear, as is done in the Röhm's experiments, we might have obtained the curve shown in Fig. 3(b).

From the results obtained here, we are forced to conclude that the difference between stress-strain curves obtained in pure gliding and those obtained in the usual tensile test is not so large as described in Röhm's paper.

[1] F. Röhm and A. Kochendörfer, Z. Metallkunde **41**, 265 (1950).
[2] W. T. Read, Pittsburgh Conference on Plastic Deformation of Crystalline Solids, p. 111, 1951 (discussion to Dehlinger's paper).

Analysis of Fluorescent X-Radiation by Means of Proportional Counters

ROBERT E. ROWLAND
Argonne National Laboratory, Lemont, Illinois
(Received March 9, 1953)

RECENTLY, several articles have appeared on the production and use of characteristic fluorescent radiation in microradiography.[1,2] This radiation has been produced by allowing the beam from an x-ray tube to fall upon an external target. The proportions of fluorescent and scattered radiations produced in this manner have usually been determined by aluminum filtration.

The purpose of this letter is to describe a method which we have utilized to analyze rapidly, and with sufficient resolution, the x-ray spectrum originating from these targets. The beam to be analyzed is intercepted through a 0.005-in. Be window by a proportional counter (4 in. wide, 12 in. long) filled at one atmosphere with 90 percent argon, 10 percent methane. This counter can detect with good efficiency x-rays between 2.5 and 15 kev.[3,4] It has been our practice to record the output of such a detector by means of a motor-driven single-channel analyzer and to register the results directly on a Brown potentiometer.

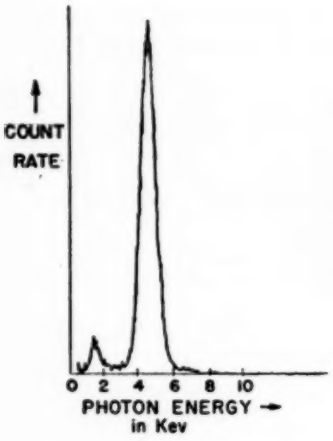


FIG. 1. Motor-driven single-channel analyzer trace of the fluorescent beam from 0.007-in. titanium. The major peak is due to the 4.52-kev K_α x-ray; the small peak at lower energy (1.53 kev) is due to the escape of the argon K_α x-ray (2.99 kev) from the counter volume.

Applied to the study of fluorescent beams, the counter has been particularly valuable in evaluating the energy and hence the source of spurious x-rays. In this manner we have determined that both the primary and the fluorescent beam must be well collimated, if the counter (or film and sample in microradiography) is to avoid seeing extraneous fluorescent material; moreover, collimating materials must be either of the same Z as the external target or of $Z \leq 15$, if monochromacity is to be retained. Aluminum should be theoretically adequate, but in practice commercial aluminum alloys were found to emit iron K_α radiation of undesirable intensity. Lucite is now employed quite satisfactorily for all collimators and support materials.

With this type of collimation a spectrum of the type shown in Fig. 1 was obtained with a Ti radiator. If the purity of the beam is defined as the number of photons within the K lines to the total number of photons in the secondary beam, then the purity of the Ti beam as a function of the peak voltage applied to a self-rectified tube with a tungsten anode and a 1-mm Be window was found to be: 95 percent at 9PKV; 94 percent at 11.5 PKV; 91 percent at 16 PKV; and 87.5 percent at 23 PKV. As the voltage applied to the x-ray tube increases, the L_α line of tungsten (8.4 kev) becomes more prominent, as can be seen in Fig. 2.

In most of our work to date, we have employed elements emitting characteristic K or L radiation between 3 and 7 kev; that is, lines whose energy has been lower than the characteristic L_α radiation (8.4 kev) of tungsten. In these cases, we find that spurious x-rays of energy greater than the line we are trying to excite are present in the "fluorescent" beam. Their intensity, however, is reduced by operating the x-ray tube at a potential which is only two or three times the energy of the line to be excited. This effect is evident in Fig. 2.

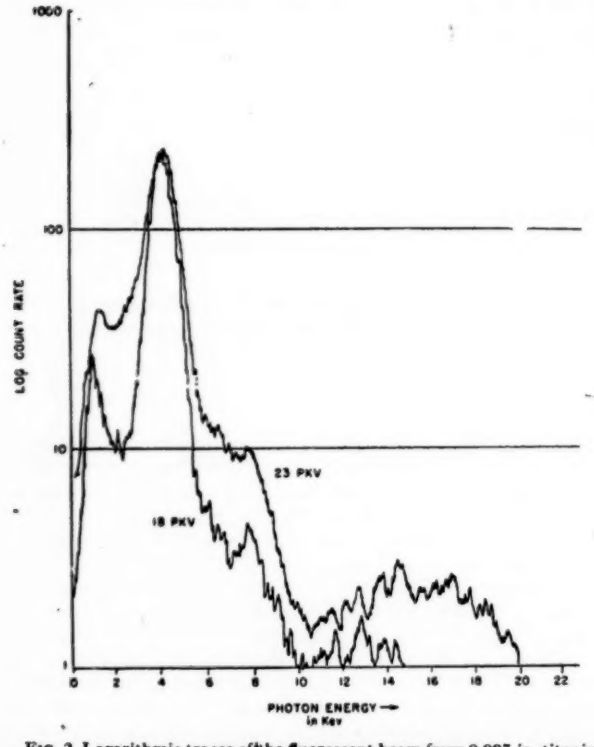


FIG. 2. Logarithmic traces of the fluorescent beam from 0.007-in. titanium radiator, showing the relative increase in the number of x-ray photons harder than the Ti K_α line as the potential applied to the x-ray tube is increased. The two curves were normalized at the Ti K_α peak.

[1] T. H. Rogers, J. Appl. Phys. 23, 881–887 (1952).
[2] H. R. Splettstosser and H. E. Seemann, J. Appl. Phys. 23, 1217–1222 (1952).
[3] Curran, Cockcroft, and Angus, Phil. Mag. 40, 53 (1949).
[4] Bernstein, Brewer, and Robinson, Nucleonics 6, No. 2, 33 (1950).

Dynamic Compliance, Dynamic Modulus, and Equivalent Voigt and Maxwell Models for Polyisobutylene*

HERBERT LEADERMAN AND ROBERT S. MARVIN
National Bureau of Standards, Washington, D. C.
(Received March 2, 1953)

DYNAMIC and stress relaxation measurements on samples from a uniform lot of polyisobutylene, with weight-average and viscosity-average molecular weights of 1.56×10^6 and 1.35×10^6, respectively, have been obtained by several investigators participating in a cooperative program initiated by the National Bureau of Standards.[1] The averaged data, when reduced to a "standard" temperature of 25°C following the procedure of Ferry,[2] with a minor modification, were used to plot both G', the real part of the dynamic modulus, and J', the real part of the dynamic compliance, shown by the solid lines in Fig. 1.

The appearance of three fairly distinct plateaus in the J' curve, which could be reproduced roughly by a Voigt model consisting of two complete Voigt elements plus a degenerate spring and a degenerate dashpot, would lead one to expect three plateaus in the G' curve, corresponding to a Maxwell model with three complete Maxwell elements in parallel. The fact that the experimental G' curve shows only two plateaus is puzzling, and led us to carry out a numerical transformation from the six-element Voigt model which best represented the compliance data to the corresponding Maxwell model.

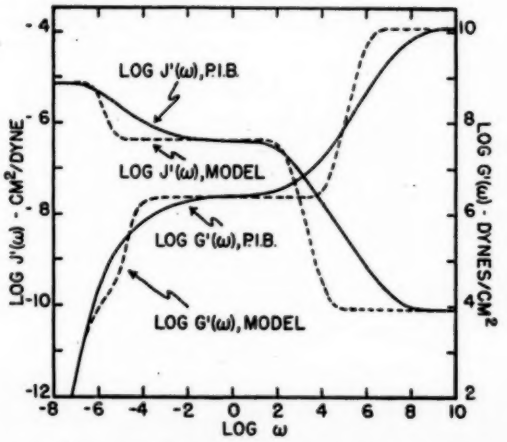


FIG. 1. Dynamic modulus and compliance of polymer and models.

The values chosen for the individual compliances in the Voigt model were obtained from the values of the J' curve at the three plateaus, the retardation times were chosen as the maxima in the logarithmic distribution function (not shown), and the value of the viscosity was that measured directly at 25°C by Ferry and Grandine,[3] and in this laboratory. They are as follows:

$$J_g = 8.3\times10^{-11}\ \text{cm}^2/\text{dyne};$$
$$J_1 = 4.0\times10^{-7}\ \text{cm}^2/\text{dyne}, \quad \tau_1 = 4.0\times10^{-3}\ \text{second};$$
$$J_2 = 7.0\times10^{-6}\ \text{cm}^2/\text{dyne}, \quad \tau_2 = 1.0\times10^{6}\ \text{seconds};$$
$$\eta = 6.16\times10^{10}\ \text{poises}.$$

For the values chosen, the parameters of the equivalent Maxwell model were calculated from the following approximate expressions, which hold for this particular case:

$$G_a = 1/J_g = 1.2\times10^{10}\ \text{dynes/cm}^2;$$
$$\tau_a = J_g\tau_1/J_1 = 8.3\times10^{-7}\ \text{second};$$
$$G_b = 1/J_1 = 2.5\times10^{6}\ \text{dynes/cm}^2;$$
$$\tau_b = J_1\tau_2\eta/(J_2\eta+\tau_2) = 1.72\times10^{4}\ \text{seconds};$$
$$G_c = \eta^2J_2/(\eta J_2+\tau_2)^2 = 1.30\times10^{4}\ \text{dynes/cm}^2;$$
$$\tau_c = \eta J_2+\tau_2 = 1.43\times10^{6}\ \text{seconds}.$$

The dotted lines in Fig. 1 represent the J' and G' curves which this particular model predicts. The three plateaus appear in the J' curve, of course, since the parameters of the Voigt model were chosen to represent this behavior. The G' curve, plotted from the parameters of the Maxwell model calculated above, shows two points of inflection in place of the third plateau which we expected, and the two long-time relaxation times are spaced so closely together that any "smearing-out" of the model behavior would be expected to mask both these inflections.

It is interesting to observe that for frequencies greater than about 10^{-5} cycle, the curves of $\log G'$ and $\log J'$ for the polymer are practically mirror images of each other. The same is only roughly true for the model curves. If the abscissa scale be considered as a scale of $\log(1/t)$, then the $\log G'$ curve represents approximately the stress relaxation behavior of the polymer or of the model.[4,5] Similarly, the $\log J'$ curve represents approximately a creep curve (with Newtonian flow subtracted) or a creep recovery curve from a steady-state condition.

By varying certain parameters of the models it is possible to reproduce changes which occur in the mechanical response curves due to a change in molecular weight or molecular weight distribution. If in the original Voigt model τ_2 and η are changed to 10^2 seconds and 6.16×10^6 poises, τ_b and τ_c are changed in the ratio of 10^4 to 1, and the lower part of the G' curve is translated four decades to the right. This corresponds to the behavior of a polymer whose weight-average molecular weight has been reduced to about 0.066 of the material for which the original model parameters were picked, and the predicted results are in agreement with those of Tobolsky and McLaughlin.[6]

If J_2 were increased from 7.0×10^{-6} to 1.4×10^{-5}, leaving the rest of the original parameters unchanged, the ratio τ_c/τ_b would be increased slightly, but no marked change in G' would result. This would correspond physically to a broadening of the molecular weight distribution without change in weight-average molecular weight,[7] and this particular calculation predicts that the G' curve should be relatively insensitive to a change in the steady-state shear compliance due to a change in molecular weight heterogenity.

It should be emphasized that the considerations presented here do not imply that the behavior of the actual polymer can be adequately represented by a six-element model. The "plateaus" in the experimental curves, for instance, are in general much less pronounced than in those calculated for the model; and if the curves for the imaginary parts of either the compliance or modulus are compared, the peaks are much sharper for the model than for the actual polymer, and the high frequency peak in G'' occurs displaced about two decades in frequency from that predicted by the calculated Maxwell model. It is, however, rather interesting to note how well the conversion from a relatively simple Voigt model to the corresponding Maxwell model can reproduce the main features of the modulus curve.

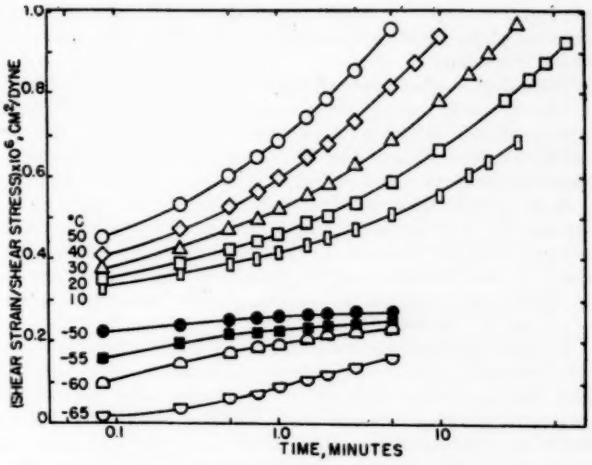


FIG. 2. Retarded elastic behavior of raw rubber.

This general shape of the compliance curve is found for other rubberlike materials for which we have creep and recovery data available. In Fig. 2 the retarded-elastic shear behavior of a specimen of raw rubber is shown.[7] Flow was negligible under the conditions of test. It is seen that there is a retardation mechanism with a time-scale of one second around −60°C, and another mechanism with the same time-scale at a temperature above 50°C. Plotted on a reduced frequency scale against $\log(1/t)$ these data would yield a single curve much like the J' of Fig. 1 between $\log\omega = -4$ and $\log\omega = +4$. The parallel-plate plastometer data of van Rossem and van der Meijden,[8] when analyzed by the procedure of Dienes,[9] are in agreement with this picture. Some recent measurements in this Bureau[10] on the "slow" retarded elastic mechanism in polyisobutylene specimens of weight average molecular weight between 1.1×10^4 and 8.5×10^4 indicate that the "average" retardation time of this "slow" mechanism is around one second when the viscosity is 10^5 poises.

Most of the data on viscoelastic properties of noncross-linked polymers in the rubberlike range published recently have been presented in terms of Maxwell distribution functions. The comparison between the J' and G' curves of Fig. 1 indicates the desirability of considering such properties also in terms of a Voigt presentation, and the advantages which are gained from measurements of creep and recovery.

* This work was supported in part by the U. S. Office of Naval Research and the U. S. Office of the Quartermaster General.

[1] R. S. Marvin, Paper to be presented at Second International Congress on Rheology, Oxford, England, July, 1953.
[3] J. D. Ferry, J. Am. Chem. Soc. 72, 3746 (1950).
[3] J. D. Ferry and L. D. Grandine, private communication.
[4] R. S. Marvin, Phys. Rev. 86, 644 (1952).
[5] A. V. Tobolsky, J. Am. Chem. Soc. 74, 3786 (1952).
[6] A. V. Tobolsky and J. R. McLaughlin, J. Polymer Sci. 8, 543 (1952).
[7] H. Leaderman, unpublished data.
[8] A. van Rossem and H. van der Meijden, Rubber Age (New York) 23, 438 (1928).
[9] G. J. Dienes, J. Colloid Sci. 2, 131 (1947).
[10] H. Leaderman and R. G. Smith, Phys. Rev. 81, 303 (1951).

X-Ray Line Broadening from Precipitation in Cu-Fe Alloys

T. S. HUTCHISON
Department of Physics, Royal Military College, Kingston, Canada
(Received January 13, 1953)

DURING the course of a study of the effects of mechanical and heat treatments on the physical properties of alloys containing small amounts of iron in copper, x-ray diffraction

photographs have revealed evidence of precipitation strain, possibly plastic in nature, existing in the copper after annealing and slow cooling.

The alloys studied contained between 0.4 and 1.34 percent by weight of iron in copper. These were originally in the hard-drawn wire condition. Filings were prepared and annealed in vacuo at 850°C for three hours. The filings were then furnace-cooled to room temperature over a period of 12 hours.

X-ray powder pictures were made, using the fillings, and long exposures of 40 hours were given in an attempt to obtain diffraction patterns from the iron. In each case the copper (400) diffraction line was microphotometered and the $K\alpha$ doublet resolved by the graphical method of Rachinger.[1] The "half-height width" of the α_1 line was then found.

The equilibrium room temperature solubility of iron in copper is extremely low,[2] but there is reason to suppose that the alloys after furnace cooling from 850°C retain about 0.26 percent iron in supersaturated solid solution.[3-5] The iron should therefore be largely in the precipitated form after heat treatment. This was confirmed by observation of the three most intense iron lines, *viz.*, (110), (211), and (200), in the diffraction photograph of the 1.34 percent Fe alloy. The (110) and (211) lines were visible in photographs of all the alloys down to that containing 0.5 percent Fe. The iron therefore precipitates in the body-centered cubic form and, from the observation that the (110) and (211) lines did not appear to be broadened, probably exists in particles of the order of 10^{-5} cm in size.

The observations of the "half-height widths" of the copper (400)$K\alpha_1$ lines indicate that an approximately linear relation exists between the width and the percentage of iron. This is shown in Fig. 1. If this graph is extrapolated back to zero iron percentage

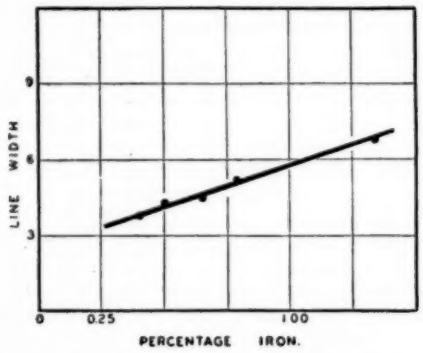


FIG. 1. Relation between the percentage iron in the Cu-Fe alloy and the width at half-height of the copper (400)$K\alpha_1$ line.

it gives approximately the width of the (400)α_1 line found with a pure copper specimen. We can thus compute a "broadening produced by the iron" without, at this stage, attempting to formulate a mechanism. The method of determining the broadening due to Jones is used.[6] The "strain" or percentage change in lattice spacing $\delta d/d$ is calculated from the broadening, and the results are shown in Fig. 2. It is seen that the strain produced by the iron is surprisingly high, reaching a value of at least one-third that found by Brindley to be caused by filing of copper.[7]

The x-ray diffraction pictures of the filed powders must certainly contain some composition broadening, owing to the inadequate homogenizing of the specimens. Suppose, in the case of an alloy containing say 1.13 percent Fe, the filings irradiated by the x-rays contain *all* percentages of iron in solution up to the solubility limit, i.e., up to 0.26 percent, for the treatment given. The lattice parameter would then have a range of values giving rise to a broadened (400) Cu line. However, lattice parameter determinations previously carried out in this laboratory give 3.6079A for pure copper and 3.6071A for an alloy containing 0.26 percent Fe. The percentage strain $\delta d/d$ associated with composition

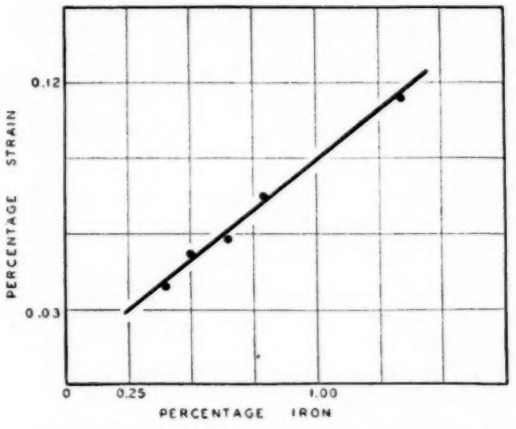


FIG. 2. Relation between the percentage strain $\delta d/d$ caused by the iron precipitate and the amount of iron present in the alloy.

broadening would then be equal to -0.025 percent. This value of $\delta d/d = -0.025$ percent is to be compared with ± 0.10 percent actually found in a 1.13 percent iron alloy. Thus composition broadening cannot account for the majority of the "x-ray strain" found.

The data are insufficient to specify exactly the cause of the line broadening found. Of the several reasons for diffuseness in reflections from such alloys, such as stacking faults, coherence of the precipitate with the matrix and so on, it might be considered, in view of the magnitude of the effect caused by the iron, that stresses in the copper have given rise to "block" deformation and thus to the x-ray broadening. Such plastic deformation, caused during precipitation, has been previously reported in the literature.[8,9] This possible explanation would then have to be completed by assuming that lattice recovery of the copper is almost certainly lessened by the presence of the precipitated particles of iron.

[1] W. A. Rachinger, J. Sci. Instr. **25**, 254 (1948).
[2] M. Hansen, *Der Aufbau der Zweistofflegierungen* (Springer, Berlin, 1936).
[3] D. Hanson and Grace W. Ford, J. Inst. Metals **32**, 335 (1924).
[4] F. E. Hetherington and J. Reekie, J. Appl. Phys. **22**, 1293 (1951).
[5] T. S. Hutchison and J. Reekie, Phys. Rev. **83**, 854 (1951).
[6] F. W. Jones, Proc. Roy. Soc. (London) **A166**, 6 (1938).
[7] G. W. Brindley and P. Ridley, Proc. Phys. Soc. (London) **50**, 501 (1938).
[8] Barrett, Kaiser, and Mehl, Trans. Am. Inst. Mining Met. Engrs. **117**, 39 (1935).
[9] A. Guinier and P. Jacquet, Rev. met. **41**, 1 (1944).

The Use of the Mercury Cathode in the Dielectric Breakdown of Liquids

A. H. SHARBAUGH, J. K. BRAGG, AND R. W. CROWE
General Electric Research Laboratory, Schenectady, New York
(Received April 6, 1953)

THE mercury electrode has often been used in studies of the dielectric breakdown of solids. It has the obvious advantage that its surface makes intimate contact with the surface of the solid dielectric at every point. This being the case, the surface of a mercury electrode takes the shape of the dielectric being studied. Although the mercury electrode is liquid, the breakdown system has the rigidity of the dielectric itself.

To use a mercury electrode with a liquid insulator, however, one has to overcome the lack of rigidity of the system; the mercury-insulator interface will move or deform under the influence of the electric field. On the other hand, there is the advantage that such an interface will be molecularly smooth if it is carefully prepared, in contrast to the inevitable roughness of even the most carefully polished solid metal surface. Points or edges on an electrode surface are undesirable because there will be

enhancement of the electric field in their neighborhoods. This may bring about breakdown, directly or through field emission induced by the enhanced field.

Another obvious advantage of the mercury electrode is the ease and speed with which it may be prepared. We use a small glass cup which is filled with mercury to the brim, so that one may obtain a convex surface. Electrical contact is made through the bottom of the cup.

Because of these marked advantages, we have devised a procedure for the use of a mercury cathode together with a conventional solid anode in breakdown studies on liquid dielectrics. We use rectangular pulses of voltage which are sufficiently short (about a microsecond) so that the surface of the mercury cathode is only slightly deformed in the duration of a pulse. The pulses, of course, must at the same time be long enough so that the measured electric strength (in the absence of electrode deformation) is independent of pulse length.

The effect of the small amount of deformation which does occur on the measured breakdown voltage can be corrected for by an extrapolation procedure. This involves the measurement of the electric strength as a function of pulse duration, as shown for benzene in Fig. 1. In the absence of electrode deformation, the

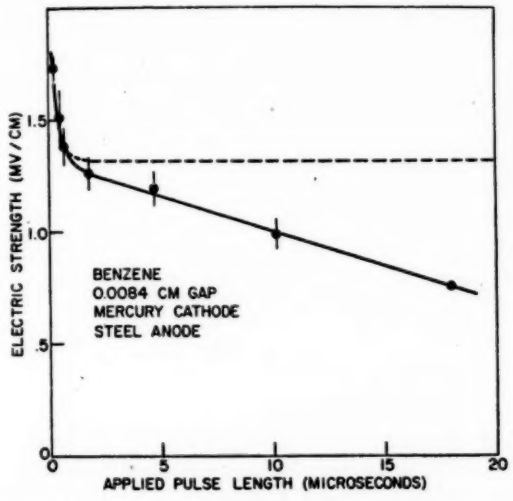


FIG. 1.

portion of the curve beyond pulse lengths of about one microsecond is known to be independent of pulse length. Electrode deformation, however, causes a gradual decrease in the apparent electric strength as the pulse duration is increased. From the slope of the curve beyond one microsecond, it is clear that the electrode deformation per microsecond is small and that the error introduced by a linear extrapolation of the measurements in this region to zero time is probably not a serious one.

Since a good deal of labor is involved in doing breakdown experiments with pulses of several different lengths, we have developed a way of correcting results which are measured at a single pulse length. The force upon the surface of a conductor in contact with a dielectric of dielectric constant ϵ, exerted by an electric field E, is proportional to ϵE^2. Now it is not at all apparent why curves such as Fig. 1 should be linear, in view of the complicated dynamical force system in operation, but all low dielectric constant ($\epsilon \leqslant 3$) liquids we have measured do indeed give straight lines in the pulse length range from one to about twenty microseconds. Furthermore, the extent of deformation after a specified length of time, as determined from data like that in Fig. 1, has turned out to be proportional to ϵE^2 for a large number of liquid dielectrics. Thus one must do a careful series of experiments on a single liquid, similar to those we have done with benzene, in order to establish the proportionality factor; thereafter measured results may be corrected from a knowledge of the observed (apparent) electric strength and the dielectric constant. The viscosity of the dielectric appears to have nothing whatever to do with the extent of deformation.

Our use of the mercury cathode over a period of time has shown us that exceedingly reproducible results are thereby obtainable. For example, one series of nineteen runs gave a standard deviation of four percent. This is presumably due to the absence of surface irregularities.

A Note on "Metallic Shadow-Casting Using a Nozzle System"

SHIGETO YAMAGUCHI
Scientific Research Institute, Ltd., Hongo, Tokyo, Japan
(Received March 9, 1953)

THE present author recognized by means of electron diffraction that a thin film obtained on a glass plate by vacuum evaporation of pure titanium held in a tungsten basket was composed of tungsten as well as of titanium. In the explanation of the electron microscopic images which have been shadow cast by evaporation of some metals held in a tungsten basket, T. Hibi neglects evaporation of tungsten.[1] A micrograph shadowed with titanium which is shown in Hibi's paper seems not to be characteristic of pure titanium. This will be the case with metals evaporated or sublimed with difficulty, e.g., Th, Mo, Pt, etc. Since the diffusion rates in mixed gases are given by a function of atomic or molecular weights (Graham's law), the metallic film prepared by vacuum evaporation through a nozzle might mainly be composed of tungsten under some conditions. The information on the shadowing qualities of elements given by Hibi is therefore not reliable for some metals.

[1] T. Hibi, J. Appl. Phys. **23**, 957 (1952).

The Expansion of a Fissionable Gas as a Source of Atomic Power

EDWARD H. KERNER*
Wayne University, Detroit, Michigan
(Received October 21, 1952)

LITTLE, if any, attention appears to have been paid to the possible use of gaseous fissionable materials in an engine of internal combustion type, wherein the compressed and at-least-critical gas cools itself by expansion and performs external work. The technical problems in the design of such a device would not be light and will not be entered into here, but the possibilities of relatively small size and high thermal efficiencies for it, as well as the elimination in it of the formidable heat-transfer problems of conventional reactors, may warrant some discussion, if only to indicate a perhaps fresh direction of thought in the field of nuclear power technology.

Mechanical power from a fissionable gas may be obtained in at least two ways, *viz.*, in a turbine or in a reciprocator. Figure 1 is a

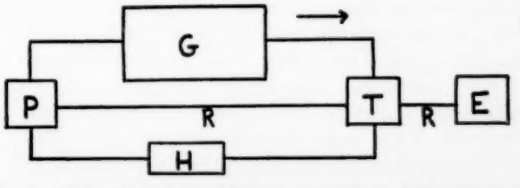


FIG. 1. Schematic diagram showing expansion of a fissionable gas into a turbine. Arrow indicates direction of gas flow.

schematic diagram of a closed system containing a fissionable gas, only one part of which, G, is a tank of such size and shape as to constitute a critical pile when its gaseous contents is at a suitable pressure. In series with G are a turbine T, pump P, and heat exchanger H. T is driven by the hot expanding gas from G, and in turn, through the rotor R, drives P and delivers work to the external apparatus E. H serves to heat the fissionable material if it has condensed, or cool it before its entry into P if it is too hot. P must pump the gas up to a sufficiently high pressure to return it to G. This is to say that a gaseous pile is inserted into a conventional turbine system so that the working gas is its own, quasi-permanent fuel by virtue of the geometry of the pile through which it passes continuously and is a fissile part.

The likely fissionable gas is uranium hexafluoride enriched in U^{235}. The technology of this corrosive chemical appears to be sufficiently advanced to permit its use in a pile; it may, for example, be safely handled with materials like nickel, monel, copper, aluminum, and dry glass; the "fluorolubes" are adequate lubricants.[1] Further fluorine and/or fluorocarbon gas could be added to stabilize the UF_6 and to moderate neutrons; solid grids of moderating material could also be introduced.

In the reciprocator, the mechanical power is obtainable by squeezing the gas briefly to supercriticality with a piston, whereupon the gas swiftly heats up and then cools itself by blowing the piston out and becoming subcritical. It is a sort of Diesel engine in short. The question is, can a suitable amount of energy be released without excessively sensitive adjustments? The following very crude analysis, addressed to this question alone, suggests that the answer is yes.

Assume that a gas of n mole of $U^{235}F_6$ is enclosed in a cylinder of base area $A=\pi r_0^2$ and variable height x; that only slow and prompt neutrons are emitted and absorbed; that their density is governed by the diffusion equation; that there is no reflector for the neutrons; that the pressure of the gas is the equilibrium pressure given by the ideal gas law and that this pressure exerts the only force on the movable piston head; and that no heat is lost from the cylinder. These assumptions effectively emphasize the possibility of violent time behavior of the gaseous pile. The neutron density N is given by $\partial N/\partial t=D\nabla^2N+KN$, where D and K are diffusion and source parameters. To avoid the mathematical difficulty of dealing with a moving boundary we make a quasi-static approximation, which should be fairly adequate for the interesting part of the piston motion that occurs in the vicinity of the piston's turning point; we assume that the instantaneous rate of energy release $\dot{Q}$ (proportional to the instantaneous average neutron density) may be found from the diffusion equation with fixed boundaries corresponding to the current value of x. We have then from the simplest solution of the diffusion equation $\dot{Q}=\dot{Q}_0e^{\zeta(x)t}$. If we use Newton's law for the piston motion, $F=PA=M\ddot{x}=(nRT)/x$; now eliminating T by using energy conservation, $nC_v\Delta T=\Delta Q-\Delta(\frac{1}{2}M\dot{x}^2)$, gives finally the differential equation

$$x(d\dot{x}/dt)+(1+R/C_v)\dot{x}\dot{x}=(R/C_v)(\dot{Q}_0/M)e^{\zeta(x)t}$$

for the piston motion. A third-order equation is needed because all three of x, $\dot{x}$, and $\ddot{x}(\sim P)$ are to be initially specified. The boundary conditions on N being taken as $N=0$ at a distance beyond the cylinder boundary equal to the "augmentation distance" δ,[2] it is found that $\zeta=\gamma/x-\alpha x$. Here γ and α depend on x [through their dependence on $\delta=\delta(x)$], but insensitively enough so that they can be regarded as fixed. Criticality occurs for $x=x_0=(\gamma/\alpha)^{\frac{1}{2}}$.

Using the known neutron cross sections and arbitrarily taking $x_0\simeq 4r_0$, a reasonable form for our model of the gaseous pile is comprised in the data: $n\simeq 10$ mole, $\gamma\simeq 5.67\ 10^5$ cm/sec, $\alpha\simeq 70$ $\text{cm}^{-1}\,\text{sec}^{-1}$, $x_0\simeq 90$ cm. The differential equation for the piston motion is difficult to integrate; however, a quick insight into the dependence of the fission heat released on the incident piston velocity is gained by calculating the "adiabatic motion" in which $\dot{Q}$ is placed equal to zero on the right side of the differential equation. With $C_v=3R$ (about right for polyatomic gases), the integration is elementary. We find an upper limit, $Q_{\max}$, to the heat released in the actual motion by means of $Q_{\max}=\int$ (over adiabatic motion) $\dot{Q}_0e^{\zeta}dt$, since if the piston were constrained to move with adiabatic motion into and out of the supercritical gas, more heat would be released than in the actual motion because the piston penetrates deeper into the gas and keeps it supercritical for a longer time. Table I illustrates what happens when $\dot{Q}_0=\frac{1}{4}$ watt, $M=75$ kg, and $\ddot{x}_0=20$ cm/cs^2 ($P_0\simeq 10$ atmospheres) for various $\dot{x}_0$; our unit of time is 1 cs$=0.01$ second, and x_1 is the turning point in the adiabatic motion. If all of $Q_{\max}$ in the last entry of the table were found as piston kinetic energy, the piston velocity would be roughly the velocity of sound. The sensitivity of $Q_{\max}$ to $\dot{x}_0$ is not impossibly large; some study of the actual motion shows this sensitivity to be quite a bit smaller.

TABLE I.

$\dot{x}_0$ cm/cs	$Q_{\max}$ kilocal	x_0-x_1 cm	$t(x_0-x_1)$ cs
18	2.58	7.66	0.834
18.5	6.5	8.06	0.853
19	18.2	8.47	0.872
19.5	52	8.91	0.891
20	152	9.33	0.910
20.5	490	9.78	0.928
21	1060	10.24	0.946

The effect of delayed neutrons will naturally modify the motion when the gas first gets critical, but very shortly criticality would be achieved on prompt neutrons alone; the delayed neutrons would delay the "explosion" in the cylinder. The control of the reciprocator could be enhanced through proper automatic cyclical manipulation of neutron absorbers, and through the cyclical application of suitable external forces to the piston.

* Present address, Physics Department, Iowa State College, Ames, Iowa.

[1] J. J. Katz and E. Rabinowitch, *The Chemistry of Uranium* (McGraw-Hill Book Company, Inc., New York, 1951), National Nuclear Energy Series.

[2] C. D. Goodman, Editor, *The Science and Engineering of Nuclear Power* (Addison-Wesley Press, Inc., Cambridge, 1947), Chap. 3.

A Nonreciprocal Microwave Component

M. L. KALES, H. N. CHAIT, AND N. G. SAKIOTIS
Antenna Research Branch, Radio Division I, Naval Research Laboratory, Washington 25, D. C.

(Received March 30, 1953)

IN a series of articles published in 1948–1949, Tellegen[1] discussed the utility of a nonreciprocal circuit element which he called a gyrator. C. L. Hogan[2] has shown that such a circuit element can be realized at microwave frequencies by utilizing the ferromagnetic Faraday rotation in ferrites.

More recently other ways of constructing nonreciprocal elements have been found at the Naval Research Laboratory. As in the case of the Faraday rotation, these depend on the fact that, in the presence of a static magnetic field, the permeability of a ferrite is a nonsymmetrical tensor. As a result, if a rectangular wave guide (Fig. 1) is partially filled with ferrite in an asymmetrical

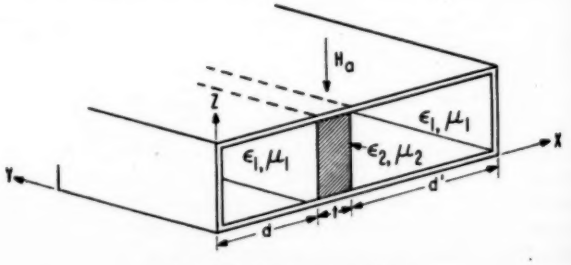


FIG. 1. Ferrite in wave guide.

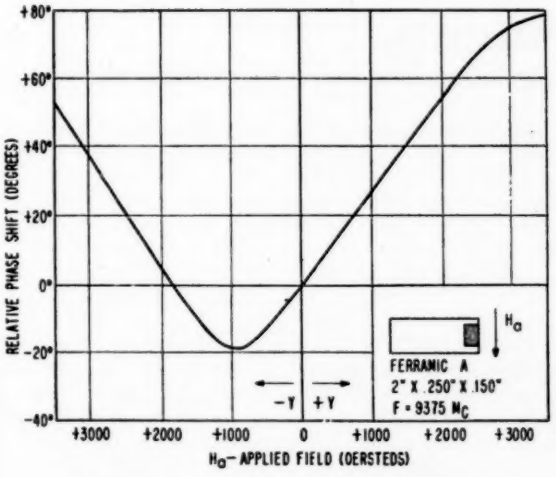


FIG. 2. Phase shift *vs* applied transverse magnetic field.

manner, and a static magnetic field H_a is applied transversely, the propagation constants of the modes will depend on the direction of propagation. This can be shown theoretically for those modes which are independent of the z coordinate.

When H_a is in the direction of the z axis, the rf components of B and H are related by the equations[3] $B_x=\mu H_x-j\mu' H_y$, $B_y=j\mu' H_x+\mu H_y$, $B_z=\mu_z H_z$. Let ϵ_2 be the dielectric constant of the ferrite, and let $\mu_2=\mu\{1-(\mu'/\mu)^2\}$. The part of the wave guide not containing the ferrite will be characterized by the constants ϵ_1 and μ_1. If an $e^{-j\beta y}$ y-dependence is assumed, then for modes which are independent of z, the propagation constant β satisfies the equation

$$\tan K_2 d[\mu_2^2 K_1^2-\mu_1^2\{K_1^2+(\beta\mu'/\mu)^2\}\tan K_1 d\tan K_1 d'] +\mu_1\mu_2 K_1 K_2[\tan K_1 d+\tan K_1 d'] -\beta\mu'/\mu\tan K_2 d[\tan K_1 d-\tan K_1 d']=0,$$

where $K_1^2=\omega^2\epsilon_1\mu_1-\beta^2$, and $K_2^2=\omega^2\epsilon_2\mu_2-\beta^2$. Since the last term of this equation contains the first power of β as a factor, whereas β is squared wherever else it appears, it is clear that in general the sign of β cannot be changed without changing magnitude of β. In other words, the magnitude of the propagation constant depends on the direction of propagation. Note that if the ferrite is centered within the guide, $d=d'$, and the last term of the equation vanishes. Thus, in this case the magnitude of the propagation constant is independent of the direction of propagation.

Figure 2 shows the measured phase shift for both directions of propagation as a function of the applied transverse field when a piece of commercial ferrite material is placed within a standard X-band wave guide.

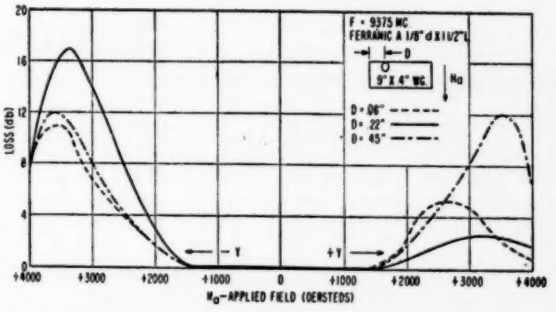


FIG. 3. Power loss *vs* applied transverse magnetic field.

It is to be expected that the attenuation constant as well as the phase constant will depend on the direction of propagation. This suggests the possibility of having low attenuation for one direction of propagation, and high attenuation for the other, and thus obtaining a one-way transmission line. Qualitatively, it can be reasoned that such a result might be obtained if the ferrite is located in the part of the guide where the H vector of the normal wave-guide mode is circularly polarized. The reason for this choice is that the sense of rotation of the H vector reverses with the direction of propagation, but ferromagnetic resonance absorption will occur for only one of the two senses of rotation. This result has been confirmed experimentally. In Fig. 3 the measured insertion loss for both directions of propagation is plotted as a function of applied H_a for several positions of ferrite. The difference in attenuation for the two directions is greatest when the ferrite is located where the H vector of the normal wave-guide mode is circularly polarized.

[1] B. D. H. Tellegen, Philips Research Reports 3, 81–101 (1948) 3; 321–337 (1948); 4, 31–37 (1949); 4, 366–369 (1949).
[2] C. L. Hogan, Bell System Tech. J. 31, No. 1, 1–31 (January, 1952).
[3] D. Polder, Phil. Mag. 40, 99–115 (January, 1949).

Errata: Emission Decay Phenomena due to the Contaminated Anode

[J. Appl. Phys. 24, 107–109 (1953)]

K. AMAKASU AND T. IMAI
The Electrical Communication Laboratory, Tokyo, Japan

IN Fig. 4 the unit of emission current should be μa not ma. The sentence from sixth line on page 109 should read "Accordingly, we consider the critical energy of 5.0 ev to be the dissociation energy of barium-oxide molecule rather than the exciting energy of oxygen ion of barium-oxide crystal."

Announcements

Fourth Symposium on Plasticity

The **Fourth Symposium on Plasticity** will be held at **Brown University, Providence, Rhode Island** on September 1, 2, and 3, 1953. The three principal topics of this Symposium will be

(1) stress-strain relations (perfectly plastic, work-hardening, visco-elastic, and visco-plastic materials),

(2) distribution of stress and strain in bodies exhibiting inelastic properties (static and dynamic problems, infinitesimal and finite deformations),

(3) structural analysis and design in the inelastic range (limit analysis, static and dynamic loading, shakedown, inelastic buckling).

Both foreign and American experts in the field will present papers covering their most recent experimental and analytical research, and everyone is invited to participate in the discussion.

Certain sessions of the Symposium, containing papers of particular interest to civil engineers, are being sponsored, jointly with Brown University, by the Engineering Mechanics Division of the American Society of Civil Engineers. The remaining sessions are under the joint sponsorship of the U. S. Office of Naval Research and Brown University.

All persons interested in receiving further information concerning the Symposium should write to Professor H. J. Weiss at the Graduate Division of Applied Mathematics, Brown University, Providence 12, Rhode Island. Copies of the program and reservation cards will be sent out approximately six weeks before the date of the Symposium.

Booklets and Bulletins

Evaluating Research and Development. Annotated Proceedings of a Conference of Research Administrators held on the campus of the University of California, Los Angeles, May 10, 1952. Editors: I. R. Weschler and Paula Brown.

National Bureau of Standards Circular 535. An annotated bibliography of selected references on the solid-state reactions of the uranium oxides.

Low Temperature Bibliography for the Field of Cryogenics. Copies of this booklet are available free of charge to who are interested in cryogenics. Write to C. L. Jewett, Mechanical Division, Arthur D. Little, Inc., 30 Memorial Drive, Cambridge 42, Massachusetts.

Consolidated Engineering Bulletin CEC-1800B. Copies may be obtained by writing to Consolidated Engineering Corporation, 300 North Sierra Madre Villa, Pasadena 15, California.

Eastman Organic Chemicals (List No. 38). Obtainable on request from the Eastman Organic Chemicals Department, Distillation Products Industries (Division of Eastman Kodak Company) Rochester 3, New York.

Meetings

Techniques for Electron Metallography, a symposium devoted to a discussion of techniques for the application of the electron microscope to metallographic problems, will be held June 30 at the **annual meeting of the American Society for Testing Materials** in Atlantic City, New Jersey. Invited and contributed papers will be presented. The symposium is being sponsored by ASTM Committee E4 on metallography.

The **American Society of Heating and Ventilating Engineers** will hold its 1953 **Semi-Annual Meeting** in Denver, June 29–July 1, at the Shirley-Savoy Hotel. Fourteen papers will be presented at four technical sessions devoted to determination of heat and moisture transfer through building materials, cooling tower performance, heat pump design and performance, effect of relative humidity on heat losses of individuals in various temperatures, and heat exchanges in floor panel heated rooms and heat flow analysis in panel heating and cooling sections.

The **First Annual Meeting** of the **Operations Research Society of America** was held in Cleveland on May 15th and 16th. The meeting featured a day-long symposium devoted to the general question, "What is Operations Research," at which the philosophy, definition, and applications of operations research to industrial problems were discussed. More specialized discussions were held at the Case Institute of Technology on May 16th.

Fellowships

Eastman Kodak Company has awarded fellowships to 17 educational institutions for the year, 1953–1954. This is part of the company's assistance to promising young scientists to encourage training in chemical research. All for advanced study, ten of the fellowships are in chemistry, four in physics, and three in chemical engineering.